Institut für Molekulare Mechanismen bei Krankheiten
der Vetsuisse-Fakultät der Universität Zürich

Institutsdirektor: Prof. Dr. med. vet. Michael O. Hottiger

Arbeit unter wissenschaftlicher Leitung von:
PD Dr. Raffaella Santoro

Epigenetic analysis of ribosomal RNA sequence heterogeneity in embryonic stem cells

Inaugural-Dissertation

zur Erlangung der Doktorwürde der
Vetsuisse-Fakultät Universität Zürich

vorgelegt von

Julia Kuhn

Tierärztin

aus Schaffhausen, Schweiz

genehmigt auf Antrag von:

PD Dr. Raffaella Santoro, Hauptreferentin
Prof. Dr. Urs Meyer, Korreferent

2017

Table of Contents

1 Zusammenfassung

Die Biogenese von Ribosomen ist ein komplexer Vorgang, dessen erster Schritt die Transkription der ribosomalen RNS (rRNS) Gene beinhaltet. Das menschliche, sowie das Maus Genom enthalten je ungefähr 400 Kopien dieser rRNS Gene. Die Gene liegen in Einheiten vor, die sich auf verschiedenen Chromosomen wiederholen. Die Transkription der rRNS Gene unterliegt epigenetischen Regulationsmechanismen. Hierbei spielt unter anderem DNS Methylierung eine wichtige Rolle. Jüngste Studien konnten zeigen, dass die epigenetische Hemmung der rRNS Gene durch den Entwicklungs- bzw. den Differenzierungsstatus einer Zelle reguliert wird. In embryonalen Stammzellen (ESCs) werden alle rRNS Gene aktiv transkribiert und erst während der Differenzierung wird die Promotorsequenz einiger Gene methyliert und diese somit in ihrer Transkription gehemmt. Dass es Stammzellen nicht möglich ist zu differenzieren wenn die Bildung von Heterochromatin in bestimmten Bereichen der rRNS Gene inhibiert wird, unterstreicht die Wichtigkeit dieses Prozesses.

Die Regulationsmechanismen, welche die Transkription von rRNS Gene während der Differenzierung orchestrieren sind größten Teils bekannt. Allerdings bleibt noch zu klären, ob die epigenetische Hemmung der rRNS Gene ein zufälliger Prozess ist (jedes rRNS Gen kann zufällig methyliert werden) oder ob bestimmte Subklassen von rRNS Genen spezifisch methyliert und so in ihrer Transkription gehemmt werden.

Frühere Studien haben gezeigt, dass rRNS Gene nicht alle die gleiche Sequenz aufweisen. Unterschiede zwischen den Sequenzen finden sich sowohl zwischen unterschiedlichen Individuen, als auch innerhalb der rRNS Sequenzen des gleichen Individuums. In der Regel unterscheiden sich die rRNS Gene besonders in ihrer “intergenic spacer region” (Sequenz zwischen den jeweiligen rRNS Genen). Oft unterscheidet sich die Längen der jeweiligen „Enhancer“ Regionen, es sind aber auch Polymorphismen einzelner Nukleotide (Single nucleotide Polymorphism SNP) bekannt. Resultate verschiedener Studien, einschließlich Ergebnisse unserer Gruppe, haben SNPs an Position -104 und +44 relativ zum Transkriptionsstartpunkt identifiziert. Folglich können die rRNS Genvarianten entsprechend dieser Resultate anhand des vorliegenden SNP unterschieden werden (-104C, -104A, +44A, +44G und +44T).

Das Ziel dieser Arbeit war es, ein System zu etablieren, welches das Quantifizieren der rRNS Genvarianten verschiedener Individuen erlaubt, sowie die Analyse deren epigenetische Merkmale ermöglicht. Im Rahmen dieses Projektes haben wir eine quantitative Polymerasen Kettenreaktion (PCR) zur Quantifizierung der SNPs entwickelt, welche zuverlässig zwischen den jeweiligen rRNS Sequenzen unterscheidet und diese in verschiedenen Zelllinien und Geweben quantitativ messen kann. Des Weiteren haben wir diese Methode so angepasst, dass wir epigenetische Modifikationen im Verlauf der Differenzierung von Stammzellen analysieren können.
Unsere Resultate zeigen eine unterschiedliche Verteilung der rRNS Genvarianten zwischen Individuen. Außerdem konnte gezeigt werden, dass die epigenetische Hemmung von rRNS Genvarianten kein zufälliger Prozess ist, sondern das gewisse rRNS Subklassen eher dazu neigen methyliert zu werden als andere. Diese Resultate legen nahe, dass bestimmte genetische Unterschiede der rRNS Sequenzen einen Einfluss auf ihren epigenetischen Status sowie ihre Transkriptionsaktivität haben. Die Entwicklung dieses spezifischen Systems wird in Zukunft sehr hilfreich sein um das Zusammenspiel von genetischen und epigenetischen Regulationsmechanismen auf die Aktivität von rRNS Genen besser zu verstehen.

2 Summary

Transcription of ribosomal RNA (rRNA) genes is the initial event of ribosome biogenesis, a complex pathway dedicated to the production of ribosomes. Human and mice genomes contain about 400 copies of rRNA genes, which are organized into tandemly repeated clusters distributed among different chromosomes. rRNA gene transcription is regulated by epigenetic-mediated mechanisms, including CpG methylation. Recent results revealed that the establishment of epigenetic silencing at rRNA genes is developmental regulated. In embryonic stem cells (ESCs) all rRNA genes are transcriptionally active and only upon differentiation a fraction of genes undergoes epigenetic silencing through methylation of the promoter sequence. Formation of heterochromatin at a fraction of rRNA genes is critical since abrogation of this process impairs ESC differentiation. Although the mechanisms by which this process is regulated during ESC differentiation are in large part understood, it still remains to clarify whether epigenetic silencing at rRNA genes is a random event (i.e. any rRNA gene can be epigenetically silenced) or whether a defined set of rRNA genes is specifically CpG methylated and transcriptionally repressed.
Early studies have shown that rRNA genes do not share the exact same sequence. Sequence heterogeneity can be observed between individuals but also between rRNA sequences in the same individual. This variation often occurs within the intergenic spacer regions and can be observed as different length of enhancer regions but also as single nucleotide polymorphism (SNP). Results from our and other laboratories have identified SNPs at position -104 and +44 relative to the transcription start site. These results indicate the presence of rRNA gene variants, which can be distinguished according to their SNP (-104C, -104A, +44A, +44G and +44T).
The aim of this work was to establish a method that allows the quantification of rRNA gene variants among individuals and determine their epigenetic and transcription state. We developed a SNP quantitative PCR, which efficiently distinguishes rRNA sequence variation and allows quantitative measurement of the amounts of rRNA gene variants present in several cell lines and tissue. Moreover, we adapted this method to analyze the epigenetic state of rRNA gene variants during embryonic stem cell differentiation. The results revealed different abundance of rRNA gene variants among individuals and determined that upon ESC differentiation rRNA gene silencing is not a random event, but

that certain rRNA gene variants are more prone to acquire CpG methylation. The results suggest that the genetic variation among rRNA sequences influences their epigenetic and transcription state. The development of this SNP specific quantitative PCR represents an important tool for future studies aimed to dissect the crosstalk between genetic and epigenetic regulation at rRNA genes, which might be useful for the understanding of complex diseases like cancer.

3 Introduction

3.1 Embryonic stem cells

Embryonic stem cells are defined by their unlimited self-renew potential and their capacity to differentiate into any of the three germ layers, that constitute the vertebrate embryo (i.e. endoderm, mesoderm and ectoderm). Hence pluripotency can give rise to all the somatic lineages of the embryo as well as to the germline (Young 2011). The development of a totipotent zygote towards a blastocyst leads to this pluripotent state in vivo. Two lineages emerge during this developmental process: the inner cell mass (ICM), which gives rise to the pluripotent cell population, and the trophectoderm (TE), which forms the extraembryonic epithelial layer that encloses and supports the ICM (Fisher and Fisher 2011). Embryonic stem cells (ESCs) can divide unlimitedly in vitro. Stem cells divide either asymmetrically or symmetrically. This process gives rise to one or two daughter stem cells with a comparable developmental potential to the mother cell. During symmetric division two identical sister cells are obtained, whereas asymmetric division gives rise to one cell, which retains the original phenotype of the mother cell, while the second cell acquire a new phenotype.

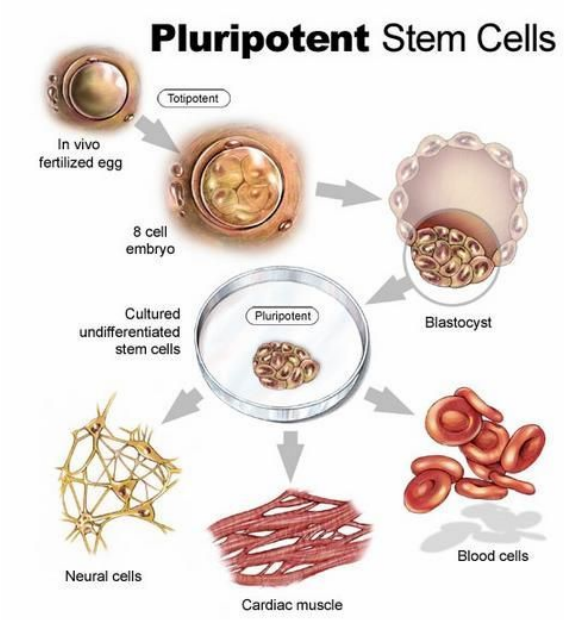

Figure 1 The pluripotency of embryonic stem cells. Schema depicts the derivation of embryonic stem cells from inner cell mass (ICM) and their potential to generate all cell types. Picture was taken from https://www.sciencelearn.org.nz/resources/1965-stem-cells.

The self-renewal capacity is essential for stem cells to expand their number during development, to maintain stem cell pool within adult tissues and to restore the stem cell pool after injury (reviewed in (He, Nakada et al. 2009)). The potency of a stem cell reflects the variety of cell types that can potentially result through differentiation. The potency ranges from the totipotency of the zygote to the unipotency of some adult stem cells such as epithelial stem cells, which enable cell turnover in epithelia (Slack 2000).
Because of these unique properties, ESCs are not only an ideal in vitro tool for studying early mammalian development, but they also offer a great therapeutic potential regarding regenerative medicine since they might be a potential source for cell and tissue replacement. In order to realize these potential clinical applications, the molecular and cellular mechanisms controlling pluripotency and differentiation need to be better understood.
ESCs are derived from the ICM of E3.5-E4.5 mouse embryos using a variety of culture conditions, such as leukemia inhibitory factor (LIF) and fetal calf serum (FCS). Mouse ESCs appear as homogenous colonies of small tightly packed cells when cultured on a feeder layer of mitotically inactivated embryonic fibroblasts as well as in the presence of fetal calf serum (FCS) (Martin 1981, Nichols, Evans et al. 1990). The efficiency of those first steps in culturing ESCs was quite low and only further adjustments of the culture conditions allowed establishing of more stable and homogenous cell lines. In 1988, two different studies by Wiliams and Smith found that the self-renewal support coming from feeder cells arose from a cytokine called leukemia inhibitory factor (LIF) produced by these cells. LIF is a member of the IL6 family that binds to the leukemia inhibitory factor receptor (LIFR) and it can be isolated from the cytokine factors produced by the feeder cells ((Smith, Heath et al. 1988),(Williams, Hilton et al. 1988)). LIF promotes ESC self-renewal by activating the transcription factor STAT3. The analysis of ESCs cultured under these conditions (LIF and serum) gave important insights into embryonic development and pluripotency.
Recent developments have enabled the derivation of ESCs in defined serum-free medium supplemented with two small-molecule kinase inhibitors (2i): PD0325901 impairs the MAP kinase pathway and therefore blocks differentiation whereas CHIR99021 activates Wnt signaling and thus enhances self-renewal (Ying, Wray et al. 2008). These ESCs are suggested to represent the ground state of pluripotency.

As compared to conventional ESCs cultured in FCS, 2i ESCs are much more homogeneous in morphology and show higher clonogenicity (≥50%) (Gertsenstein, Nutter et al. 2010). There is growing evidence that not only the signalling but also the epigenetic make-up differs, suggesting that 2i and serum ESCs represent two different states of pluripotency. Therefore, it has been considered that 2i ESCs reflect earlier embryonic stages as serum ESCs (Marks and Stunnenberg 2014).

3.1.1 Transcriptional network of pluripotency

A complex network of transcription factors that activates self-renewal related genes and represses differentiation pathways is responsible for the maintenance of pluripotency in stem cells. In the past decades, several master regulators of this network have been identified, which include Oct4, SOX2 and Nanog that are considered to be crucial for the maintenance of the undifferentiated state.
During mouse embryonic development, Oct4 is expressed within the totipotent blastomeres, pluripotent epiblast as well as in primordial germ cells (PGCs) ((Rosner, Vigano et al. 1990), (Scholer, Dressler et al. 1990)). SOX2 is expressed within the ICM and the extraembryonic ectoderm of pre-implantation embryos (Avilion, Nicolis et al. 2003). Since Oct4 or SOX2 KO embryos are not able to form a pluripotent ICM and differentiate into trophectoderm it is suggested that both factors are essential to establish and maintain pluripotency (Avilion, Nicolis et al. 2003). In the absence of the leukemia inhibitor factor (LIF), Nanog is capable to sustain mESC self-renewal. Although Nanog KO embryos do not posses a pluripotent ICM, Nanog KO ESCs can be generated afterwards. Those ESCs are still pluripotent yet they tend to differentiate. Although they cannot give rise to PGCs they are still capable to contribute to chimeras, suggesting that Nanog is indispensable to establish the pluripotency of the ICM and to safeguard the pluripotent status of the germ line ((Chambers, Silva et al. 2007), (Mitsui, Tokuzawa et al. 2003)).
The exact regulation of these transcription factors is crucial, as their over or under-expression would affect ESC identity and differentiation state. Several studies mapping the genomic-binding sites of these core transcription factors revealed that these factors co-occupy the promoters of several transcription factors regulating their expression. Especially they bind to genes involved either in maintenance of pluripotency or in

establishing differentiation programs. The interaction with these promoters leads to differential regulation. On one hand it represses the transcription of genes involved in differentiation, on the other hand it maintains the active state of pluripotency-related genes. Interestingly the core pluripotency factors also co-occupy their own promoters and therefore are enhancing their own transcription enforcing the maintenance of pluripotency and self-renewal ((Boyer, Lee et al. 2005), (Loh, Wu et al. 2006)).

3.1.2 The epigenetic state of ESC genome

Since embryonic stem cells are competent to generate any cell type, a high level of genome plasticity is needed to rapidly establish a transcriptional program, allowing the selection of a specific direction of differentiation. In contrast to somatic cells, ESCs display an open chromatin structure. This finding suggests that the ratio between euchromatin and heterochromatin is higher and that the chromatin structure is less condensed than in differentiated cells (Bhattacharya, Talwar et al. 2009). A wide range of active histone marks like trimethylation of lysine 4 of histone H3 (H3K4me3), acetylation of lysine 27 of histone H3 (H3K27ac) and acetylation of histone H4 (H4ac), which contribute to keep chromatin accessibility, are associated with the euchromatic state of ESC genome ((Azuara, Perry et al. 2006),(Gaspar-Maia, Alajem et al. 2011)). Despite the transcriptional permissive state of ESCs, in order to maintain pluripotency and genome stability, a certain degree of gene silencing must take place. For example, genes related to differentiation pathways are in a silent state due to the repressive histone mark trimethylation of lysine 27 of histone H3 (H3K27me3). Also the transcription of sequences with a high mutagenic potential like retrotransposons must indeed be repressed (Reik 2007). It is interesting to note, that some of these genes are also marked with the active histone mark H3K4me3 at their promoter and for this reason have been defined as “bivalent” genes (Voigt, Tee et al. 2013). These bivalent domains are suggested to poise expression of developmental genes hence allowing well-timed activation while maintaining repression in the absence of differentiation signals (Voigt, LeRoy et al. 2012). Therefore upon differentiation, most of the bivalent patterns of histone modifications are erased. In the case of induction of gene activation of specific lineage developmental genes, H3K4me3 is retained whereas

H3K27me3 is released. In the case of gene downregulation, H3K27me3 is retained whereas H3K4me3 is released (reviewed in (Gaspar-Maia, Alajem et al. 2011)).
In contrast to developmental genes, which need to be loosely repressed, transposable elements for example must be completely and tightly repressed in order to avoid harmful genomic recombination events. This is achieved by H3K9 methylation and also by CpG methylation ((Bourc'his and Bestor 2004),(Schlesinger and Goff 2015)).

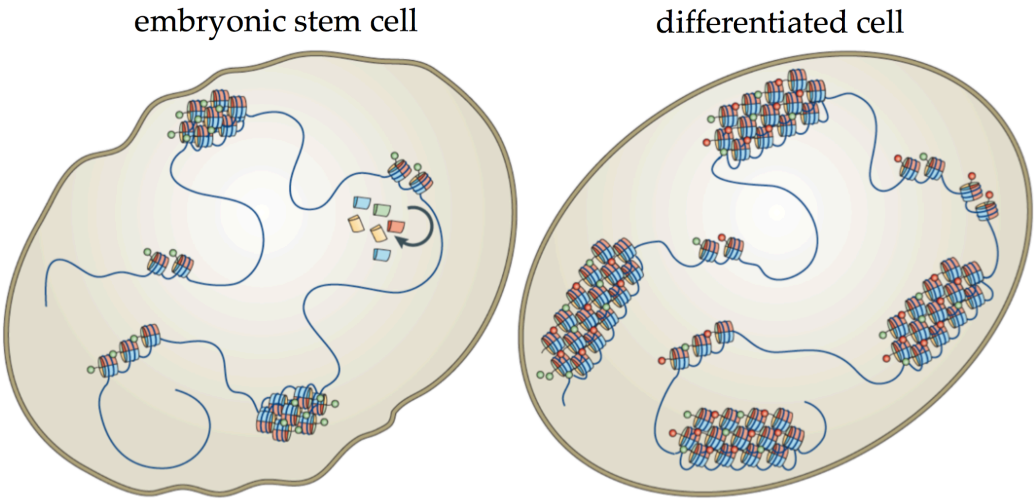

Figure 2 Chromatin during embryonic stem cell differentiation. In pluripotent embryonic stem cells (left), chromatin is globally decondensed and enriched in active histone marks (green circular tags). Upon differentiation (right), large blocks of condensed heterochromatic region form and repressive histone marks (red circular tags). Image from (Meshorer and Misteli 2006)

Stem cell differentiation involve a major rearrangement of the epigenome that leads to the repression of pluripotency genes and activation of specific subsets of developmental genes according to the differentiation program that is activated (Meissner 2010). Noteworthy, ESC chromatin undergoes structural remodelling towards a highly condensed heterochromatic and transcriptionally repressed form upon differentiation (Bhattacharya, Talwar et al. 2009) (**Figure 2**). The open genome structure of embryonic stem cells clearly reflects the plasticity and the transcriptionally permissiveness of the ESC genome that has to have the ability to enter any distinct transcriptional program for lineage specification ((Garneau, Dupuis et al. 2010),(Gorkin, Leung et al. 2014)).

3.2 DNA methylation

Like histone methylation, DNA methylation plays a role in regulating gene expression of cells. It is a highly conserved epigenetic modification of the DNA, which occurs in prokaryotes as well as in eukaryotes. In higher eukaryotes DNA methylation generally occurs at 5'-position in a cytosine-phosphate-guanine (CpG) dinucleotide (Jones 2012). Large parts of vertebrate genome contain few CpG dinucleotides, which tend to be converted to 5-methylcytosines (5mC). Unmethylated CpGs are often found organized in clusters called CpG islands (CGIs) (reviewed in (Smith and Meissner 2013)). DNA methylation can change the structure of the DNA and can alter the binding of DNA binding proteins, which may therefore modulate transcription. It is considered as a mark of silent, inactive chromatin and is inversely correlated with gene activity (Jones 1999). Mammals use the methylation of CpG-rich promoters to prevent initiation of transcription and to ensure the silencing of genes for example on the inactive X chromosome, imprinted genes and parasitic DNAs (Jones and Takai 2001). In somatic tissues, CpG methylation shows global patterns based on relative CpG density: CpG islands at housekeeping or developmental promoters are largely unmethylated, whereas non-regulatory CpGs distributed somewhere else in the genome are mostly methylated, which therefore prevents transcription (Doerks, Copley et al. 2002). The acquisition of a genome wide program of CpG methylation is essential to determine the exit of the pluripotent state of ESCs and to start differentiation ((Gifford, Ziller et al. 2013),(Mohn and Schubeler 2009)). Cytosine methylation is carried out by a group of enzymes called DNA methyltransferases (DNMT) catalysing the transfer of a methyl group from S-adenosyl-methionine to CpG dinucleotides ((Tsai, Manor et al. 2010), (Jeltsch 2006)) (**Figure 3**). Methylation patterns are established *de novo* during embryonic development and stably inherited upon cell division.

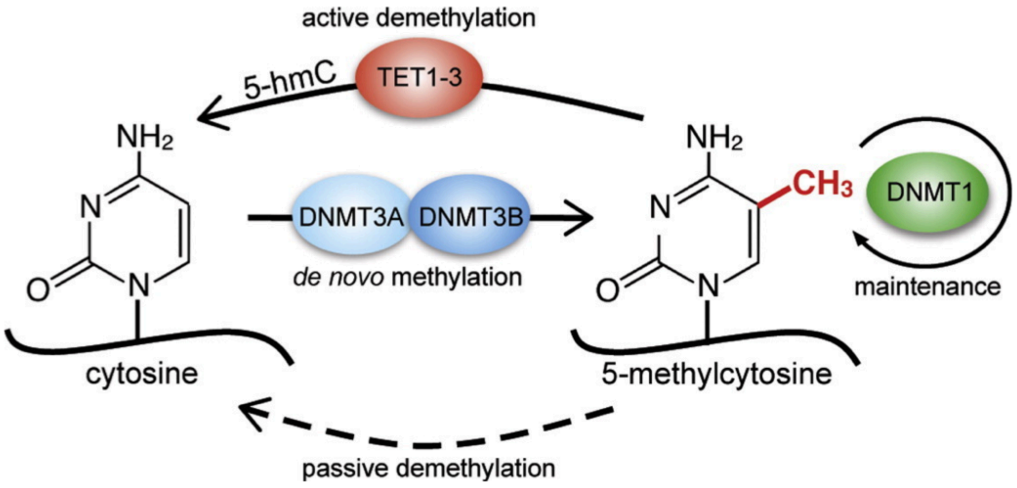

Figure 3 Biogenesis of DNA methylation in eukaryotes. DNA methylation predominantly occurs at the fifth carbon atom of cytosine bases. Methylation is catalyzed by the *de novo* DNA methyltransferases DNMT3A and DNMT3B. During replication, After the passage of the replication fork, the maintenance DNMT, DNMT1, reestablishes CpG methylation on the newly synthesized strand hence allowing maintenance of the DNA methylation pattern over generations. Passive DNA demethylation is considered to be achieved across cell division in the absence of DNMT1 maintenance activity. The mammalian TET1–3 proteins have been condidered to play a role in active demethylation since they convert 5-mC to its oxidised derivative 5-hydroxymethylcytosine (5-hmC) and further to 5-formylcytosine and 5-carboxylcytosine. These modifications are removed through DNA repair processes or are passively lost through replication. Image from (Ambrosi, Manzo et al. 2017)

DNMT1 has been known as the enzyme responsible for maintenance of CpG methylation after replication (Bestor 2000). After the passage of the replication fork DNMT1 re-establishes CpG methylation on the newly synthesized strand hence allowing maintenance of the DNA methylation pattern over generations (Hermann, Goyal et al. 2004). *De novo* establishment of DNA methylation is performed by DNMT3a and DNMT3b, which are able to act on hemi-methylated but preferentially on unmethylated CpG-dinucleotides (Yokochi and Robertson 2002). It is known that all the DNMTs are essential for correct development in mouse and in human. Using homozygous mutants the role of DNMTs in ESCs has been analysed. De novo methylation in ESCs and mouse embryos is blocked through inactivation of genes encoding *Dnmt3a* and *Dnmt3b*. A double knockout of these genes is embryonic lethal in mouse, single KO leads to embryonic (*Dnmt3b*-KO) or postnatal (*Dnmt3a*-KO) lethality (Okano, Bell et al. 1999). Interestingly, in the same work *Dnmt3a* and *Dnmt3b* KO (dKO) ESCs have been obtained but they displayed absence of de novo methylation activity, suggesting that *de novo* methylation is actually needed later than blastocyst stage during embryonic development. Mutation of *Dnmt1* gene results in embryonic lethality (Li, Bestor et al. 1992). Moreover

Dnmt1 KO ESCs maintain pluripotency but show elevated mutation rates and die shortly after being induced to differentiate ((Chen, Pettersson et al. 1998),(Jackson, Johnson et al. 2004)). These and other studies imply the importance of DNA methylation for maintaining genome stability, establishing and maintaining stable cellular identities, silencing transposable elements, genomic imprinting and X-chromosome inactivation (Wu and Zhang 2010).

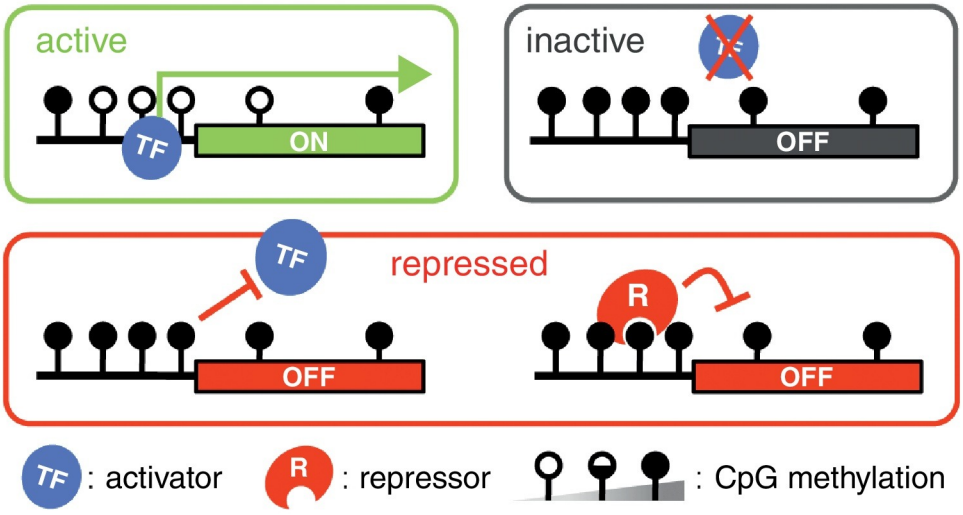

Figure 4 DNA methylation of active and repressed genes. Binding of activators and/or increased CpG density protects promoters of active genes from DNA methylation. The lack of activators leads to DNA methylation at promoters of inactive genes. Direct or indirect interference of DNA methylation with activators at promoter regions of inactive genes leads to repression. This can occur for example via mCpG-dependent recruitment of repressors to methylated sites. Image from (Baubec and Schubeler 2014)

DNA methylation affects gene expression either via direct or indirect mechanisms (**Figure 4**). Some DNA binding proteins, as for example certain transcription factors are only able to interact with their target sequence if unmethylated. Therefore CpG methylation directly abolishes their interaction with the DNA and leads to lower transcription levels (Clark, Harrison et al. 1997). Furthermore methylated DNA can be recognized by a set of proteins called methyl-CpG-binding proteins (MBPs). It has been shown, that further factors, which mediate transcriptional repression such as histone deacetylases are recruited by these MBPs, leading to the establishment of silent chromatin at methylated CpG sequences (Defossez and Stancheva 2011). To ensure heritable change in chromatin structure the DNA methylation can be copied after DNA syntheses in somatic cells, although this process seems to be highly dynamic during

embryogenesis. The global DNA methylation is gradually reduced after fertilization until the implantation at the early blastocyst stage. The average level of CpG methylation is about 20% in early blastocyst cells of the ICM3.5. A major wave of DNA methylation occurs after implantation of the embryo. This process appears to be of importance in lineage restriction and cellular potency. At E6.5, the genome is hypermethylated with an average CpG methylation level of about ~70% ((Smith, Chan et al. 2012) and reviewed in (Marks and Stunnenberg 2014)).

3.3 Nucleolus and ribosomal RNA (rRNA) genes

The nucleolus is a well-defined, membrane-less nuclear subdomain, which appears as a dense structure in electron microscopy pictures. It is the nuclear compartment where ribosomal RNA is synthetized, processed and the assembly of ribosomes takes place. Ribosome biogenesis is initiated by the transcription of hundreds of ribosomal RNA (rRNA) genes (400 in mouse and human cells), which generate rRNA precursors (45S pre-rRNA, in mouse cells, 47S pre-rRNA in human cells). These transcripts are then later cleaved and processed into 28S, 18S and 5.8S rRNAs. To form the large and small ribosomal subunits, rRNA transcripts are assembled with ribosomal proteins (Santoro 2005). Since growing cells have an excessive demand for proteins, up to ten million ribosomes per cell are required. Therefore an average mammalian cell can produce as many as 10’000 ribosomes/min. To meet the demand for this large-scale ribosome synthesis, cells have to invest a large part of their own metabolic effort. Thus it is crucial that cells keep the transcriptional activity of rRNA genes under tight surveillance to limit excessive energy consumption that could potentially deplete the cells from nutrients required for other essential processes. This is underlined by the fact that conditions that reduce cellular metabolism downregulate rRNA transcription (Moss 2004). Since the mammalian ribosome is made up of two-thirds RNA and one-third protein, this large-scale ribosome biogenesis forces cells to provide a large amount of rRNA. Thus it is not surprising that in mammalian cells, around 35% of nuclear transcription is committed to the production of rRNAs (reviewed in (Moss, Langlois et al. 2007)).

3.3.1 Structure of rRNA genes

To meet the enormous amount of ribosomes needed in a living cell, an exclusive and potent transcription system was evolved. First by amplifying the numbers of rRNA genes to hundreds or even thousands of copies per genome and second by using a specific RNA polymerase (RNA polymerase I, Pol I). Interestingly rRNA gene units are teeming with polymerases and emerging transcript complexes, while on RNA Pol II genes rarely more than one isolated polymerase is active (Santoro 2011).
The nucleolus results from the fusion of certain nucleolar organizing regions (Morin, Johnson et al. 2010). The human and mice genome contains about 400 copies of rRNA genes and each rRNA gene unit covers roughly 43kb in human and 45kb in mouse ((Gonzalez and Sylvester 1995), (Grozdanov, Georgiev et al. 2003)) (**Figure 5**). These gene copies are organized into tandemly repeated clusters that are distributed among different chromosomes. While in mice, rRNA clusters are located within the centromeric regions of chromosomes 12, 15, 16, 18 and 19, the human rRNA genes are placed between the satellite body and the short arm of acrocentric chromosomes 13, 14, 15, 21 and 26 ((Dev, Tantravahi et al. 1977),(Kurihara, Suh et al. 1994)).

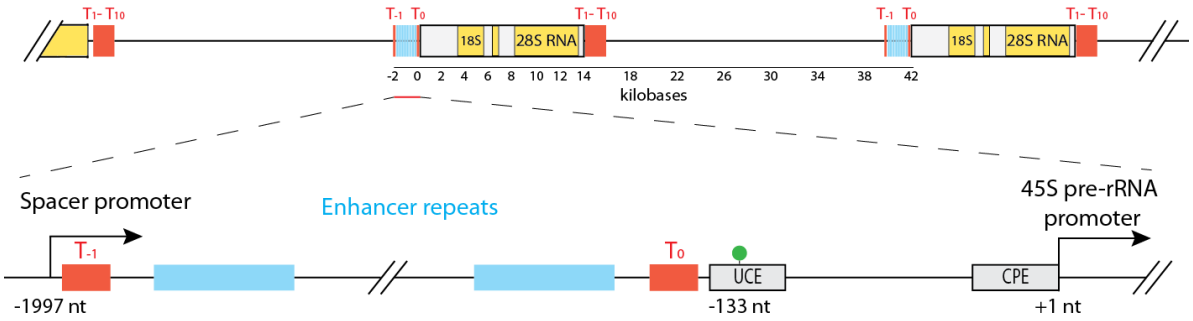

Figure 5 Organization of mouse rRNA genes. Upper panel shows the organization of one rRNA gene repeat. Scale bars (in kb) are shown below. 0 kb indicates the 5' end of the pre-rRNA transcript. Lower panel represents the magnification of the regulatory region for transcription. The transcription initiation sites of the 45S pre-rRNA and transcripts from the intergenic spacer transcripts are shown with black arrows. Red boxes indicate terminator elements (T). The blue boxes exemplify repetitive enhancer elements located between the spacer promoter and the transcription start site (TSS) proximal core promoter element (CPE). The upstream control element (UCE) is also shown and is located upstream the TSS. The green dot represents the CpG at position -133 which is crucial for rDNA silencing through DNA methylation. Adapted from (McStay and Grummt 2008).

Long intergenic spacers (IGS) of roughly 30kb are separating one pre-rRNA gene from the next (13-14kb). The IGS sequence contains regulatory elements, including the main gene promoter, an alternative spacer promoter, repetitive enhancer elements and

transcription terminators. The rRNA gene promoter contains two structure elements: the core promoter element bordering the transcription start site and roughly 100 nucleotides further upstream the upstream control element (UCE) ((Haltiner, Smale et al. 1986), (Learned, Learned et al. 1986)). One or more terminator elements that are recognized by TTF1 (transcription termination factor) are flanking the mammalian rRNA gene transcription units at their 5' and 3' ends. TTF1 is a specific DNA binding protein that stops the elongating Pol I and plays an important role in epigenetic regulation of rRNA genes. It appears that the major part of the IGS is devoid of regulatory elements, instead it comprises a high density of simple sequence repeats and transposable elements (reviewed in (Sylvester, Gonzales et al. 2004)).

To transcribe rRNA genes, the formation of the pre-initiation complex on the promoter is necessary. This pre-initiation complex comprises the binding of the upstream binding factor (UBF) and the promoter selectivity factor (SL1 in humans, TIFIB in mice) ((Clos, Buttgereit et al. 1986), (Grummt 2003), (Moss, Langlois et al. 2007)). UBF stimulates the RNA pol I promoter escape and promotes transcriptional elongation and is therefore counteracting the repressive function of the heterochromatin protein 1 (HP1) at the rRNA genes ((Panov, Friedrich et al. 2006), (Kuhn and Grummt 1992), (Pelletier, Stefanovsky et al. 2000), (Stefanovsky, Langlois et al. 2006)). Thus UBF contributes in activating and promoting rRNA transcription through all these functions. Furthermore UBF binding to rRNA genes is impaired through methylation of the CpG at the position -133 inside the UCE resulting in transcriptional repression (Santoro and Grummt 2001).

SL1/TIFIB is indispensable for the promoter specificity and contains the TATA box binding protein (TBP) and at least three Pol I- specific TBP-associated factors (TAFIs), TAFI110/95, TAFI68 and TAFI48 ((Comai, Tanese et al. 1992), (Zomerdijk, Beckmann et al. 1994), (Heix, Zomerdijk et al. 1997)). TAFIs play an important role in the assembly of the transcription complex, mediating specific interactions between the rRNA gene promoter and Pol I. They interact with UBF and by binding TIFIA they recruit Pol I to the rRNA gene. TIFIA is a basal regulatory factor that is associated with the initiation-competent subpopulation of Pol I.

At the 3' end of each rRNA gene the terminators (T) are found in ten copies (T1-T10). They are involved in termination of transcription and in determining the direction of the

replication fork over the rRNA sequences. The T sequences are bound by TTF1 ((Grummt, Rosenbauer et al. 1986), (Gerber, Gogel et al. 1997))
Between the spacer promoter and the rRNA gene main promoter several repeated sequences are located that act as enhancer of rRNA transcription and are also bound by UBF (Pikaard, Pape et al. 1990). It has been shown that different variants of rRNA units exist according to the number of enhancer repeats (in mouse 6, 9, 10, 11, 12 and 22). The variation in rRNA sequences will be discussed in more details in chapter **3.3.6**.

3.3.2 Chromatin organization of rRNA genes

Although cells contain many rRNA gene copies, not all the rRNA genes transcribe. Miller spreading analyses have shown two classes of rRNA genes: active genes, which are covered by elongating polymerases that synthesize rRNA and therefore show the characteristic tree-like appearance (DNA “trunk” from which closely packed ribonucleoprotein “branches” of increasing length extend) and silent genes, which are not associated with Pol I and hence are not transcribed. Therefore researchers proposed two ways how cells can regulate rRNA levels: (1) by controlling the transcription rate per gene by acting directly on the Pol I transcription machinery; (2) by regulating the number of genes to be transcribed (Santoro 2011)
Psoralen is an intercalating drug that leads to DNA crosslink at sites that are not protected by nucleosomes. Recent studies using psoralen have shown that rRNA genes exist in two specific forms. rRNA genes which contain nucleosomes are not accessible to psoralen and thus represent inactive gene copies, whereas rRNA genes with a chromatin structure free of regularly spaced nucleosomes are accessible to psoralen and are therefore representing active gene copies, which are transcribed ((Lucchini and Sogo 1992), (Conconi, Widmer et al. 1989)). The relative amount of these two chromatin states is similar in both growing and resting cells as well as during interphase and metaphase, suggesting that these two chromatin structures are stably propagated throughout the cell cycle and are maintained independently of the transcriptional activity (reviewed in (Santoro 2005)).
Studies demonstrated that also different epigenetic marks characterize the silent and active rRNA genes. Silent genes (inaccessible to psoralen) show more CpG methylation,

which is associated with heritable gene silencing and leads to a more heterochromatic structure. Active genes (rRNA genes accessible to psoralen) on the other hand miss this epigenetic mark, which suggests that rRNA gene methylation and repression of transcription are intimately linked (Stancheva, Lucchini et al. 1997). Later studies demonstrated a direct role of DNA methylation in repressing rRNA gene transcription. Treating mouse cells with 5-azacytidine, an inhibitor of cytosine methylation, increased pre-rRNA synthesis by 40-50%, suggesting that the lack of DNA methylation alleviates transcriptional repression of silent rRNA genes (Santoro 2005). A few crucial CpGs within the main rRNA gene promoter region seem to be critical for the repressive action of DNA methylation on rRNA transcription. Methylation of one specific CpG at -133 within the UCE (upstream control element) site impairs the binding of the Pol I transcription factor UBF (upstream binding factor) to rRNA gene chromatin. Consistent with this, methylation of one single HpaII site (CCGG) located in the rat promoter region of silent rDNA chromatin, showed a quite strong correlation with the repressed transcriptional state (Stancheva, Lucchini et al. 1997).

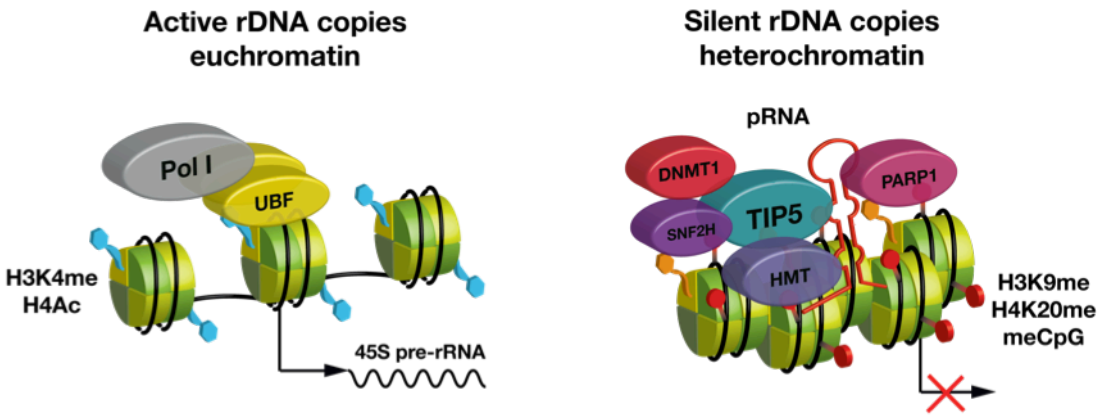

Figure 6 Model of active and silent rRNA genes. On the left side transcriptionally active rRNA genes are shown. They are characterized by open chromatin marked with active histone marks (Blue dots) and are bound by UBF and Pol I. On the right side the transcriptionally silent genes are shown. Typically they are in a compact heterochromatic structure marked with silencing histone marks (orange dots) and DNA CpG methylation (red dots). The pRNA dependent binding of TIP5 to rDNA promoters and the following recruitment of histone modifiers and DNA methyltransferases are responsible for the establishment of these repressive marks

The analysis of the composition of active and silent rRNA chromatin in higher eukaryotes and plants has been hampered for long time by the fact that rRNA genes have the same sequences. The identification of protein factors, including posttranslational modified histones, that bind either active (i.e. lack of meCpG) or to silent (i.e. enriched in meCpG) genes was only possible after the development of an assay based on chromatin immunoprecipitation (ChIP) combined with CpG methylation measurement (ChIP-chop) (Santoro, Li et al. 2002). Using this approach different studies found that the promoter of mouse and human active and unmethylated rRNA genes is associated with Pol I transcription factors and active histone marks (i.e. H4Ac and H3K4me2) (Santoro, Li et al. 2002, Santoro and Grummt 2005) whereas silent and CpG methylated rRNA genes are associated with silent histone marks like H3K9me2, H3K27me3, H4K20me3 and the heterochromatin protein 1 (HP1) (Santoro, Li et al. 2002, Santoro and Grummt 2005) (**Figure 6**).

3.3.3 The nucleolar remodelling complex (NoRC)

The nucleolar remodelling complex (NoRC) is the key player in establishing and maintaining the heterochromatic state of silent rRNA genes. The complex consists of TIP5 (TTF1-interacting protein 5) and the ATPase SNF2h ((Santoro, Li et al. 2002), (Li, Santoro et al. 2005),(Strohner, Nemeth et al. 2001), (Zhou, Santoro et al. 2002)) (**Figure 6**). SNF2h is the human homolog of the Drosophila ISWI whose ATPase activity is crucial for nucleosome remodelling. TIP5 was found in a yeast two-hybrid screening for TTF1 interacting proteins (Strohner, Nemeth et al. 2001). TIP5, also known as BAZ2A is a member of the bromodomain adjacent zinc finger (BAZ) protein family (Jones, Hamana et al. 2000). It is the largest subunit of NoRC and shares a number of important domains with other components of nucleosome remodelling complexes such as ACF, WCRF, CHRAC and WICH ((Ito, Levenstein et al. 1999), (Bochar, Savard et al. 2000),(LeRoy, Loyola et al. 2000), (Bozhenok, Wade et al. 2002)).
A bromodomain at the C-terminus allows the binding to acetylated histones and is required to mediate HDAC1 recruitment, SNF2h and other factors like HMTs and DNMTs bin a plant homeodomain (PHD) and a TAM (TIP5/ARBD/MBD) domain mediates the binding of RNA, in particular the stem loop structure of the lncRNA pRNA ((Zhou, Santoro et al. 2002),(Zhou and Grummt 2005), (Mayer, Schmitz et al. 2006)). There are further domains including a WAKZ motif, BAZ1 and BAZ2 motifs and several AT-hooks motifs. The recruitment of NoRC to the nucleoli is mediated by the transcription termination factor 1 (TT1) ((Nemeth, Strohner et al. 2004, Mayer, Neubert et al. 2008), (Savic, Bar et al. 2014)) that binds a sequence motif flaking the 5’ and 3’ ends of the rRNA gene transcription units (Olson 2011). Later the NoRC complex recruits the SIN3 co-repressor complex in a TIP5 dependent way leading to the deacetylation of nucleosomes at rRNA genes (Zhou, Santoro et al. 2002).

3.3.4 pRNA and the regulation of ribosomal genes

The lncRNA pRNA is crucial for the epigenetic silencing of rRNA genes (**Figure 6**). pRNA is an approximately 200 nt long lncRNA that corresponds to the rRNA main promoter sequence. It is one of the few trans-acting lncRNA that have been known so far ((Lee 2012), (Vance and Ponting 2014), (Froberg et al. 2014)). pRNA derives from the processing of intergenic spacer rRNA (IGS-rRNA). IGS-rRNA in turn is transcribed by Pol I starting at the alternative spacer promoter located around 2kb upstream of the main rRNA gene promoter and it spans the intergenic sequence including the main rRNA promoter sequences ((Mayer, Schmitz et al. 2006), (Santoro, Schmitz et al. 2010), (Savic, Bar et al. 2014)).

It was shown that knockdown of pRNA induces loss of heterochromatin at rRNA genes in differentiated cell, suggesting the crucial role of this lncRNA (Mayer, Schmitz et al. 2006). pRNA forms a conserved hairpin structure that is essential to bind the TAM domain of TIP5 and to recruit the NoRC complex to rDNA promoters ((Mayer, Schmitz et al. 2006), (Savic, Bar et al. 2014)).

Moreover the interaction of TIP5 with regulatory factors including PARP1 and TTF1 is mediated by pRNA, which acts as a scaffold through the stem loop structure. Moreover, pRNA acts as a guide for TIP5 to rRNA genes by promoting the association with TTF1 ((Mayer, Neubert et al. 2008), (Guetg, Scheifele et al. 2012), (Savic, Bar et al. 2014)). Recent studies have shown that mutating pRNA sequences involved in the stem loop structure formation abolishes TIP5-TTF1 interaction and therefore impairs the recruitment of TIP5 to rRNA genes (Savic, Bar et al. 2014).

3.3.5 Establishment of heterochromatin at rRNA genes

Recent studies have shown that unlike in somatic cells all rRNA genes in ESCs are in an active state and devoid of heterochromatic marks (**Figure 7**) (Savic, Bar et al. 2014). This observation is remarkable as it is well known that the reduction of heterochromatin at rRNA genes in differentiated cells leads to genomic instability due to recombination events between the repetitive sequences ((Guetg, Lienemann et al. 2010), (Peng and Karpen 2007), (Straight, Shou et al. 1999)). Yet it is still unknown how stem cells can handle such a high risk of deleterious recombination events, which also include centromeric repeats and retrotransposons.

It has been shown that the impairment of TIP5 association with rRNA genes is responsible for the lack of heterochromatin at rRNA genes in ESCs (**Figure 7**). Only upon differentiation TIP5 re-localizes within the nucleoli, where it binds and silences rRNA genes (Savic, Bar et al. 2014). The determinant for the impairment of TIP5 association and silencing of RNA genes in ESC is the lack of the processing of IGS-rRNA into the mature pRNA, a reaction mediated by the RNA helicase DHX9 (Leone, Bar et al. 2017)). In ESCs, the processing of IGS-rRNA into pRNA is impaired and activated only upon differentiation. Transfection of mature pRNA into ESCs was sufficient for the recruitment of TIP5 and formation of heterochromatin at rRNA genes, suggesting that maturation of IGS-rRNA into pRNA is crucial for the formation of heterochromatin in the nucleolus. Further experiments revealed that only the mature form of pRNA allows the association of TIP5 with TTF1 at rRNA gene promoter, whereas binding of TIP5 with the unprocessed IGS-rRNA transcript impairs this process (Savic, Bar et al. 2014). Therefore this study suggests that the recruitment of TIP5 to the rRNA genes occurs via a protein-RNA-protein-DNA module.

The formation of heterochromatin at rRNA genes appears to have a function that goes beyond the regulation of rRNA synthesis and ribosome biogenesis. The addition of mature pRNA in ESCs not only lead to rRNA gene silencing but also established highly condensed chromatin structures outside of the nucleolus, reminding of the genome organization that characterizes differentiated cells (Savic, Bar et al. 2014). This was accompanied by an increase of global levels of H3K9me2 as well as at the

heterochromatic repetitive sequences, such as major and minor satellites, as found in differentiated cells. Moreover, such heterochromatic ESCs were primed for differentiation due to upregulation of genes involved in differentiation and developmental processes (Savic, Bar et al. 2014). Hence these results suggest that the nucleolus is not only the cellular compartment of the ribosomal production but also of the heterochromatin establishment and is therefore affecting the genome architecture of the rest of the nucleus. Recent studies have revealed that the formation of heterochromatin at rRNA genes is required for ESC differentiation. Depletion of TIP5 or abrogation of IGS-rRNA processing through knockdown of DHX9 impairs ESC differentiation (Leone, Bar et al. 2017)). Remarkably, the differentiation defects of DHX9 depleted ESCs could be reverted by the addition of pRNA, whereas providing IGS-rRNA and pRNA mutants deficient for TIP5 binding are not sufficient.

Taken together these results highlight an important function linked to the heterochromatic state of rRNA genes in orchestrating the remodeling of ESC genome from an open to a compacted structure and that this process is required to exit from pluripotency.

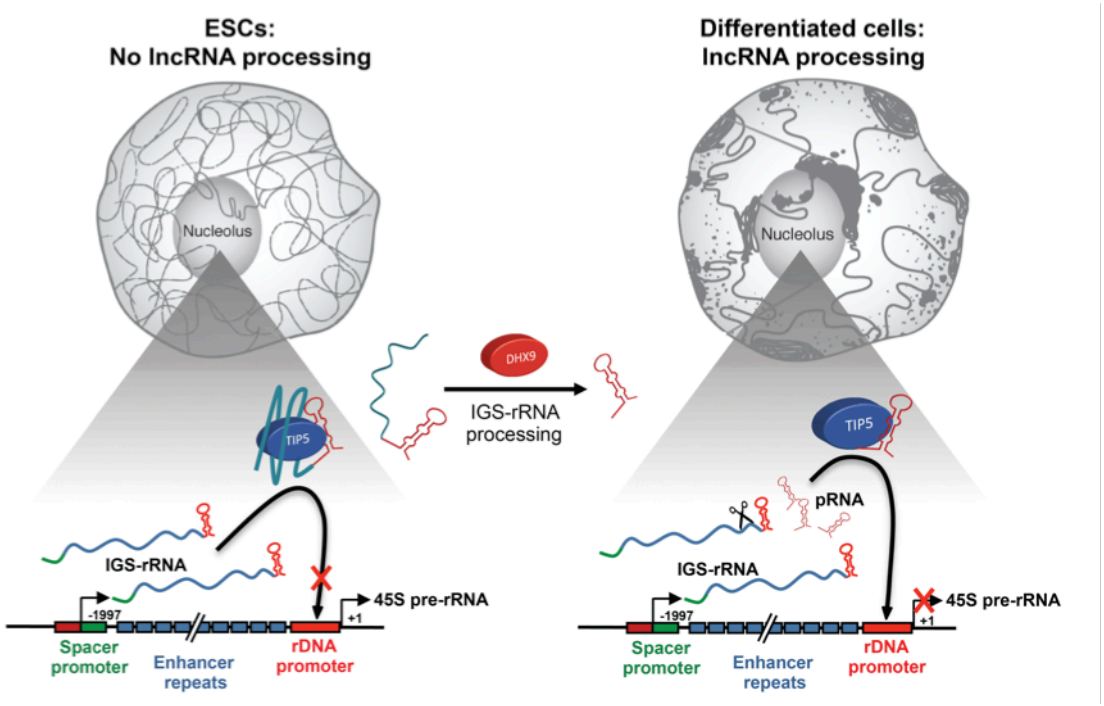

Figure 7 Model of the establishment of heterochromatin at rRNA gene during ESC differentiation. The establishment of heterochromatin at rRNA genes depends on the processing of IGS-rRNA into pRNA, a reaction mediated by DHX9 that is impaired in embryonic stem cells (ESCs) and activated only upon differentiation. The production of mature pRNA is essential since it guides the repressor TIP5 to rRNA genes, and IGS-rRNA abolishes this process. The cell cartoons represent the chromatin state of ESCs (open and euchromatic) and differentiated cells (closed and heterochromatic). The dark grey regions showed the presence of heterochromatic cluster characterizing differentiated cells.

3.3.6 rRNA gene variants

As already described in **3.3.1**, rRNA genes are clustered as tandem repeat units of roughly 44 kbp length, which are distributed among different chromosomes (in mouse cells, chromosomes 12, 15, 16, 18 and 19). In situ hybridization and cytogenetic studies showed that not all individual chromosomal rRNA gene loci are equally active in various cell types, leading to the suggestion that certain regulatory sub-domains might exist in the rRNA gene array and that they are regulated in a cell-type-specific manner (reviewed in (Tseng, Chou et al. 2008)). Early studies have shown that not all rRNA genes share the same sequences. Enzymatic digestion with the restriction enzymes EcoRI and HindIII showed heterogeneous patterns of restriction fragments in the non-transcribed spacer DNA among individuals such as BALB/c, C57BL and *Mus poschiavinus* strains and human leukemic blood cells (Arnheim and Southern 1977). Such a heterogeneous patterns of restriction fragments is clearly indicative of sequence variations among individuals. Importantly, analysis with restriction endonucleases EcoRI, HindIII and BamHI of rRNA genes in BALB/c strain revealed two, and possibly four, classes of repeating units and it was suggested that the rRNA gene heterogeneity within one individual probably resides in the very large non-transcribed spacer region (Cory and Adams 1977). A similar variation in sequence was also found in the spacer rRNA region from human placenta and it was suggested that unequal homologous exchange is the molecular basis for the observed length heterogeneity in the spacer rRNA sequences and may be a common mechanism for the generation of human genetic diversity (Erickson and Schmickel 1985). Studies in X*enopus laevis* rRNA gene heterogeneity revealed variability in the copy numbers of a short repeated sequence of two regions within the non-transcribed spacer DNA (Wellauer, Dawid et al. 1976). In humans this variable region was observed by hybridizing Southern blots of BamHI digested genomic DNA with a probe specific for the 3'-end of the 28S rRNA sequences. The presence of BamHI fragments of 6.0, 6.7, 7.6 and 8.5 kb length is characterizing this variation. Interestingly the relative distribution of fragments within one size class is characteristic for individuals. Furthermore identical patterns of fragment size and distribution have been shown in different tissues from the same individual (reviewed in (Erickson and Schmickel 1985)). It has been suggested that

recombination between ribosomal genes might be responsible for both the sequence homogeneity of the transcribed gene and the length heterogeneity in the spacer region (Smith 1976). This restriction fragment length polymorphism (RFLP) was associated with a variable number of enhancer repeats located between the spacer promoter and the promoter-proximal terminator ((Arnheim and Kuehn 1979), (Tseng, Chou et al. 2008)). Comparison of these variants by cloning and sequencing indicated that mouse spacer and promoter sequences are identical but it is the number of enhancer repeats that varies (9,10,11,12 and 22 repeats) (Santoro, Schmitz et al. 2010). Remarkably, this work found evidences for the existence of regulatory sub-domains in the rRNA gene array. The analysis of IGS-rRNA length by RT-PCR has shown that IGS-rRNA is transcribed from a specific subclass of rRNA genes containing nine enhancer repeats (Santoro, Schmitz et al. 2010). As described in **3.3.5** chapter, transcription of IGS-RNA and maturation of pRNA is required for NoRC-dependent heterochromatin formation and transcriptional silencing. Therefore these findings suggest that a specific subclass of rRNA genes is involved in regulating gene silencing and nucleolar organization. Remarkably, recent studies have shown that rRNA genes can vary within the same individual not only for the length in the intergenic spacer but also for the presence of single nucleotide polymorphisms (SNPs). For example, a certain genetic variation has been described at position -104, where either a C or an A can be found (Shiao, Leighty et al. 2011). Interestingly, these SNPs appear to influence rRNA gene expression. Indeed, protein restricted diet during pregnancy lead to a specific CpG methylation at position -133 of -104A-rRNA variant in the offspring, which correlates negatively with weaning weight (Holland, Lowe et al. 2016). Finally, a recent analysis revealed that in mouse cells rRNA genes contain a SNP (A or G or T) at +44 relative to transcription start site (Guetg, Lienemann et al. 2010). These studies taken together revealed inter and intra-heterogeneity among rRNA gene sequences between individuals. However, the mechanisms by which this sequence variation influences gene transcription remain elusive.

4 Aims of the project

As described in the previous section, establishment of epigenetic silencing at rRNA genes is developmental regulated. In embryonic stem cells (ESCs) all rRNA genes are transcriptionally active and only upon differentiation a fraction of genes undergo epigenetic silencing through methylation of the promoter sequence. Formation of heterochromatin at a fraction of rRNA genes is critical since abrogation of this process impairs ESC differentiation. Although the mechanisms by which this process is regulated during ESC differentiation are in large part understood, it still remains to clarify whether epigenetic silencing at rRNA genes is a random event (i.e. any rRNA gene can be epigenetically silenced) or whether a defined set of rRNA genes is specifically CpG methylated and transcriptionally repressed.

Several evidences have started to indicate that rRNA genes do not share the exact same sequence. Sequence heterogeneity can be observed between individuals but also between rRNA sequences in the same individuals. Results from our and other laboratories have identified single nucleotide polymorphism (SNP) at position -104 and +44 relative to the transcription start site, indicating the presence of rRNA gene variants, which can be distinguished according to their SNP (-104C, -104A, +44A, +44G and +44T).

The aim of this work was to determine whether genetic variation at rRNA genes can influence their epigenetic and transcriptional state. An important part of this project was the establishment of a method that allows the quantification of rRNA gene variants among individuals and determines their epigenetic and transcription state.

This work describes the establishment of a SNP quantitative PCR, which efficiently distinguishes rRNA sequence variation and allows quantitative measurement of the amounts of rRNA gene variants present in several cell lines and tissue. Moreover, we adapted this method to analyse the epigenetic state of rRNA gene variants during embryonic stem cell differentiation.

The results revealed different abundance of rRNA gene variants among individuals and determined that upon ESC differentiation rRNA gene silencing is not a random event but that certain rRNA gene variants are more prone to acquire CpG methylation. The results suggest that the genetic variation among rRNA sequences influences their epigenetic and transcription state. The development of this SNP specific quantitative PCR represents an

important tool for future studies aimed to dissect the crosstalk between genetic and epigenetic regulation at rRNA genes.

5 Material and Methods

5.1 Material

5.1.1 Cell lines

Name	Origin
ESC+2i	Mouse embryonic stem cells derived from *mus musculus* strain 129/Ola.
ESC+serum	Mouse embryonic stem cells derived from *mus musculus* strain 129/Ola. A kind gift for Prof. Ciaudo (ETH, Zurich)
NIH 3T3	Mouse embryonic fibroblasts

5.1.2 Medium for cell culture

Name	Composition
N2B27-medium	244 ml Neurobasal Medium 244 ml DMEM-F12 (Dulbecco's Modified Eagle Medium, Sigma) 2,5 ml N2 supplement 5 ml B27 supplement 5 ml Penicillin/Strep/Glutamine 0,5 ml Beta-Mercaptoethanol (100x, GIBCO)
FCS-medium	410 ml DMEM (Dulbecco's Modified Eagle Medium, Sigma) 50 ml fetal calf serum 10 % 5 ml Penicillin/Streptomycin /L-Glutamine (Life technologies) 1 ml Beta-Mercaptoethanol (100x, GIBCO) 5 ml Non-essential Amino Acids (100x, Life Technologies) 5 ml Sodium-Pyruvate (100x, GIBCO)

Complete medium	435 ml DMEM-F12 (Dulbecco's Modified Eagle Medium, Sigma) 50 ml 10 % fetal calf serum 5 ml Sodium-Pyruvate (100x GIBCO) 5 ml Non-essential Amino Acids (100x Life Technologies) 0,91 ml Beta-Mercaptoethanol (100x, GIBCO) 5 ml Pen/Step/Glutamine (Life technologies)
DMEM	Dulbecco's Modified Eagle Medium, Sigma

5.1.3 Buffers and solutions

Name	Composition
1x PBS buffer	140 mM NaCl 2.7 mM KCl 8.1 mM Na_2HPO_4 1.5 mM KH_2PO_4 pH 7.4
Proteinase K digestion buffer	100 mM NaCl 10 mM Tris-HCl pH8 25 mM EDTA 0,5 % SDS
DNA loading buffer	30 mM EDTA 30 % Glycerol 0.5 % Bromophenol blue

5.1.4 Chemicals, enzymes, reagents and kits

Name	Company	Further Information
Trypsin-EDTA	Gibco® Invitrogen	10x, 5 % Trypsin
Penicillin/Streptomycin	Life technologies	10'000 U/ml Penicillin, 10'000 µg/ml Streptomycin
FCS	Gibco® Invitrogen	Fetal calf serum
RNase A	Fermentas	DNase and protease-free, 10 mg/ml
Proteinase K	Fermentas	21.2 mg/ml
MgCl2	ApRoche	25 mM
Primer random dN6	Roche	2 µg/m l
dNTPs	Fermentas	dATP, dCTP, dGTP and dTTP; 100 mM
SYBR GREEN master mix	Bioline	2x SensiMix™ SYBR Kit
HpaII	New England BioLabs	10000 U/ml
NEBuffer 3.1	New England BioLabs	
Phenol:Chloroform:Isoamyl alcohol	Sigma	25:24:1
Ethanol	Merck	100 % / 70 %
Isopropanol	Merck	100 %
Agarose	Promega	
Glycogen	Roche	20 µg/µl
Natrium acetate (NaAc)	Sigma	3 M, pH 5.5
Natrium cloride (NaCl)	Sigma	3 M
Ethylenediaminetetraacetic acid (EDTA)	Sigma	0.5 M, pH 8.0
Sodium dodecyl sulfate (SDS)	Sigma	10 % (w/v)

Diethyl dicarbonate DEPC	Sigma	
Trizol	Molecular Research Center,	
DNase I	Fermentas	1 U/μl
DNase I buffer	Fermentas	
Reverse Transcriptase (RT)	Roche	50 U/μl
RT buffer	Roche	100 mM Tris, pH 8; 500 mM KCl
RiboLock RNase inhibitor	Roche	40 U/μl
Pfu DNA Polymerase	Promega	2 U/ μl
Pfu DNA Polymerase buffer	Promega	
DpnI	New England BioLabs	
DpnI digestion buffer	New England BioLabs	
Buffer RES	Macherey-Nagel	For plasmid purification with NucleoBond® Xtra Maxi/Midi column
Buffer LYS	Macherey-Nagel	For plasmid purification with NucleoBond® Xtra Maxi/Midi column
Buffer EQU	Macherey-Nagel	For plasmid purification with NucleoBond® Xtra Maxi/Midi column
Buffer NEU	Macherey-Nagel	For plasmid purification with NucleoBond® Xtra Maxi/Midi column
Buffer WASH	Macherey-Nagel	For plasmid purification with NucleoBond® Xtra Maxi/Midi column
Buffer ELU	Macherey-Nagel	For plasmid purification with NucleoBond® Xtra Maxi/Midi column
Buffer C1	Qiagen	For genomic DNA extraction with the Blood and Cell Culture DNA midi kit

Buffer RES	Macherey-Nagel	For plasmid purification with NucleoBond® Xtra Maxi/Midi column
Buffer LYS	Macherey-Nagel	For plasmid purification with NucleoBond® Xtra Maxi/Midi column

5.1.5 Equipment

Name	Company	Further Information
NanoDrop®	NanoDrop technologies, Thermo Scientific	ND-1000, Spectrophotometer
PCR machine	Applied Biosystems	GeneAmp® PCR System 2720
qPCR machine	Corbett Research	Rotor-Gene RG-3000 A

5.1.6 Primers for quantitative PCR (qPCR)

Name	5.1.6.1.1 Sequence 5' → 3'
-165/-145 forward	GACCAGTTGTTCCTTTGAGG
+111/+130 reverse	GACAGCTTCAGGCACCGCGA
-105/-87 forward	CCCAGGTATGACTTCCAG
-21/-1 reverse	ACCTATCTCCAGGTCCAATAG
+1/+20 forward	ACTGACACGCTGTCCTTTCC
+87/+66 reverse	TAGGCTGGACAAGCAAAACAG
+550/+570 forward	CTCTTGTTCTGTGTCTGCC
+745/+765 reverse	GCCCGCTGGCAGAACGAGAAG
2409-2428 pBuescript forward	CCTCCGATCGTTGTCAGAAG
2587-2606 pBuescript reverse	CGCGGTATTATCCCGTATTG
2554-2573 pBuescript reverse	GCAACTCGGTCGCCGCATAC

5.1.7 SNP primers for qPCR

Name	Sequence 5' → 3'
+44A reverse	TAAATCGAAAGGGTCTCTTT
+44G reverse	TAAATCGAAAGGGTCTCTTC
+44T reverse	TAAATCGAAAGGGTCTCTTA
-104C forward	GTCATTTTTGGGCCACCTCCC
-104C reverse	ATTACCTGGAAGTCATACCTGG
-104A forward	GTCATTTTTGGGCCACCTCCA
-104A reverse	AATACCTGGAAGTCATACCTGT

5.2 Methods

5.2.1 Cell culture

5.2.1.1 Culture conditions of embryonic stem cells

Mouse embryonic stem cells were cultured in N2B27 or FCS medium at 37 ºC. For passaging, cells were first washed with 1x PBS and subsequentially incubated with 1x Trypsin for 2-3 min at 37 ºC. Cells were then collected in inhibitor-free medium and Trypsin was removed by centrifugation at 200 x g for 5 min. After supernatant was removed the cell pellet was resuspended in respective medium and 8000 cells/cm^2 were seeded on cell culture dishes (Corning) coated with 0,1 % gelatin.

5.2.1.2 Culture conditions of neural progenitor cells

The differentiation of ESCs into neural progenitor cells (Bibel, Richter et al. 2004) was obtained using neural differentiation media (Complete medium). 27000 cells/cm^2 were seeded on a not adherent plate (Greier) preventing them from attaching. During the 8-day differentiation procedure, media was exchanged every second day. For the last 4 days of differentiation the media was supplemented with 2μl retinoic acid (RA) to generate neuronal precursours. For changing the media the plates were tilt over so the cells were sinking to the bottom and the supernatant could be removed.

5.2.2 Purification of genomic DNA

About 300’000-500’000 cells were first washed with 1x PBS and subsequentially incubated with 1x Trypsin for 2-3 min. Cells were then collected in inhibitor-free medium and centrifuged at 1000 x g for 5 min. Cell pellet was then resuspended in 300 μl lysis buffer and 1,6 μl Proteinase K (20 mg/ml) was added. Samples were incubated at 50 °C over night. After cooling the samples down to 37 °C, 1 μl /ml RNase A was added for 1 hour. Subsequently 300 μl Phenol:Chloroform:Isoamylalcohole (25:24:1) was added, samples were inverted and incubated for 5 min at room temperature. After centrifugation

at 7000 x g for 7 min at room temperature the aqueous phase was collected, 30 µl NaAC, 0.5 µl. Glycogen and 210 µl Isopropanol were added and samples were incubated for 20 min at -80 °C. Samples were then centrifuged with 20000 x g for 10 min at 4 °C, the supernatant was discarded; the pellet was washed with 500 µl Ethanol (70 %) followed by another centrifugation (20000 x g, 10 min, 4 °C). After discarding the supernatant pellets were dried at 50 °C for 3-5 min and resuspended in 40 µl water.
Amount of genomic DNA (gDNA) was measured by spectrophotometer (Nanodrop). Quality and amounts of gDNA were also analysed by agarose gel electrophoresis.

5.2.3 Enzymatic digestion of genomic DNA

Enzymatic digestions were performed in 60 µl volume containing 4 µg of genomic DNA, 5 ng pBuescript plasmid DNA, 1x Cut Smart Buffer (new England Bio Labs), 40 U HpaII (according to manufacture's protocol). Reactions were incubated at 37 ºC over night slightly shaking. 100 µl Phenol:Chloroform:Isoamylalcohole (25:24:1) was added, samples were inverted and incubated for 5 min at room temperature. After centrifugation at 7000 x g for 7 min at room temperature the aqueous phase was collected, 10 µl NaAc, 0.5 µl Glycogen and 70 µl 100 % EtOH were added and samples were incubated for 20 min at -80 °C. Samples were then centrifuged with 20000 x g for 10 min at 4 °C, the supernatant was discarded, the pellets were dried at 50 °C for 3-5 min and resuspended in 400 µl water.

5.2.4 Quantitative PCR (qPCR)

Quantitative PCR reactions were performed in 10 µl volume (1x SYBR GREEN master mix, 200 nM of respective primers). In case of RT-qPCR, quantification of gDNA was performed by amplification of 2 µl extracted/digested gDNA (10 ng/µl). Standard reactions were performed using the following thermal program: 10 min at 95 °C; 40-45 cycles of 20 s at 95 °C, 30 s at 60 °C and 30 s at 72 °C; melting from 55 or 60 °C to 99 °C with steps of +1 °C.
Quantifications were performed using standard curve, which represent serial dilutions of plasmid DNA (0.2 ng to 0.00002 ng).

5.2.5 RNA purification

First cells were washed with 1 ml 1xPBS and pelleted at 1000 g for 5 min at 4 °C. 500 µl Trizol was added and tubes were incubated at room temperature for 5 min before centrifugation at 10600 x g for 5 min at 4 °C. The supernatant was transferred to a new 1.5 ml Eppendorf, 100 µl chloroform was added and the samples were vortexed intensely for 15 s (until sample became turbid) and incubated at room temperature for another 15 min. Subsequently the samples were centrifuged at 12000 x g for 15 min at 4 °C. The aqueous phase carrying the RNA was transferred into new Eppendorf tubes containing 250 µl isopropanol. After mixing samples briefly they were placed at -80 °C for 15-20 min. For pelleting the RNA the samples were centrifuged at 13000 x g for 10 min at 4 °C. RNA pellets were then washed with 500 µl 75 % EtOH, dried and dissolved in 30 µl DEPC (diethylpyrocarbonate) treated H2O. After measuring RNA concentration by spectrophotometer (Nanodrop) samples were diluted to a concentration lower than 3000 ng/µl. Samples containing less than 3000 ng/µl were brought up to a volume of 44µl with DEPC treated H_2O to which 5 µl DNase I buffer and 1 µl of DNase I was added. The samples were incubated at 37 °C for 1 h slightly shaking. The DNase I reaction was stopped by adding 500 µl Trizol and 100 µl chloroform, followed by vortexing intensely and incubating for 15 min at room temperature. After centrifugation at 12000 x g for 15 min at 4 °C the aqueous phase was transferred to new Eppendorf tubes containing 60 µl the NaAc pH 4.8 and 300 µl isopropanol. The samples were vortexed briefly and placed at -80 °C for 15-20 min. RNA was pelleted by centrifugation at 15000 x g for 10 min at 4 °C, washed with 500 µl 70 % EtOH and dried. Finally the dried RNA pellet was dissolved in 20 µl DEPC treated H_2O.

5.2.6 Reverse Transcription

For reverser transcription the RNA samples were diluted to a concentration of 200 ng/μl and 1 μg of RNA was pipetted UltraFlux Flat Cap PCR tubes (Scientific Specialities Inc.) containing a reverse transcription mix consisting of the following TaqMan® Reverse Transcription Reagents (Roche):

- 4,4 μl 25mM $MgCl_2$
- 2 μl 10x RT buffer
- 4 μl dNTPs 2.5 mM
- 0,25 μl RNase inhibitor (40 U/μl)
- 1 μl 6mer primer 2.5 uM
- 0,5 μl Reverse Transcriptase (50 U/μl)
- DEPC treated H_2O to fill up to a volume of 15 μl

The PCR tubes were placed in the 2720 Thermal Cycler (Applied Biosystems) and the reverse transcription program was initiated (10 min at 25 °C, 60 min at 42 °C, 5 min at 95 °C).

5.2.7 Cloning

5.2.7.1 Site directed mutagenesis

Mutations in rRNA gene sequences were generated by site directed mutagenesis. PCR reactions were performed in 20 μl volume and contained 20 ng plasmid DNA, 1x Pfu Polymerase buffer, 0.8 μl of respective primers, 200 μl of dNTPs, 1 U Pfu Polymerase. Amplification were performed using the following thermal program: 5 minutes at 95 ºC; 16 cycles of 30 seconds 95 ºC, 1 minute 55 ºC, 15 minutes 72 ºC and then another 7 minutes at 72 ºC. To eliminate the original plasmids and select the mutated copies 10 U DpnI were added and samples were incubated at 37 °C for 30 min.

5.2.7.2 Transformation

To transform bacteria with DpnI digested plasmid DNA, competent bacteria cells (DH5 alpha Escherichia coli) were thawed slowly at 4 ºC. 10 μl of PCR reaction (see site directed mutagensis, 4.3.1) were added to 50 μl bacteria by gently mixing and kept a 4 °C for 10 minutes. Samples were heat-shocked for 45 seconds at 42 ºC and then placed on ice for 2 minutes. After addition of 500 μl LB medium, samples were incubated for one hour at 37 ºC, centrifuged at 1000 x g for 5 minutes. Pellet was resuspended in ca. 100 μl LB medium and plated on ampicillin agar plates over night at 37 ºC. The day after, colonies were picked and further cultured with 2 ml LB medium in the presence of ampicillin (100 μg/ml).

5.2.7.3 Miniprep

2 ml of over night-cultured bacteria were centrifuged (3000 x g, 10 minutes, 4 ºC) and bacteria pellet was resuspended in 250 μl resuspension buffer containing RNase A. Bacteria were lysed by adding 250 μl lysis buffer, gently mixed and incubated for 4 minutes. 350 μl neutralization buffer was added, samples were inverted several times, and centrifuged for 10 minutes (20000 x g) at 4 ºC. Supernatants were collected and transferred into new Eppendorf tubes containing 420 μl isopropanol. Samples were vortexed, incubated for 10 minutes at -20 ºC s and centrifuged at 20000 x g for 10 minutes at 4 ºC. Pellet was air dried and then resuspended in 50 μl dH_2O with shaking for 30 minutes at 37 ºC. To identify correct clones 1.2 μg of plasmid were sent for sequencing.

5.2.7.4 Midiprep

120 ml of bacteria, cultured overnight at 37 °C, were centrifuged at 3000 x g for 10 minutes and then resuspended with 8 ml of resuspension buffer in a 50 ml falcon. Bacteria were lysed by adding 8 ml lysis buffer. Samples were gently mixed and incubated for 5 minutes at room temperature. Reactions were neutralized by adding 8 ml of neutralization buffer by gently mixing. Lysates were poured on filters placed inside columns previously equilibrated with 12 ml equilibration buffer. Filters and columns were

then washed with 5 ml equilibration solution. After removal of filters, the column was washed with 8 ml of washing solution. Plasmid DNA eluted from column with 5 ml elution buffer was collected into a 15 ml falcon and precipitated by adding 3,5 ml isopropanol. Samples were left for 10 minutes at room temperature and then centrifuged for 45 min, 4816 x g at 4 ºC. The DNA pellet was then resuspended in 500 μl TE buffer, transferred into a 2 ml Eppendorf tube, and again precipitated by adding 50 μl of NaAc and 1 ml of 100 % Ethanol followed by vortexing, cooling for 1 h at -80 ºC and centrifugation at 20000 x g for 10 minutes at 4 ºC. Pellet was air dried and then resuspended in 200 μl TE-buffer. DNA concentration was measured by Nanodrop. Samples were then diluted to 1 μg/μl plasmid DNA with TE-buffer and stored at 4 °C.

6 Results

6.1 Establishment of a polymorphism specific quantitative PCR method to analyse rRNA gene variants

As described in **3.3.6**, accumulating evidences indicated that rRNA genes do not have the same exact sequence, showing different length in the intergenic regions such as the mouse enhancer repeat region as well as single nucleotide polymorphisms (SNPs) at position +44 (rRNA gene variants +44A/T/G) and -104 (rRNA gene variants -104C/A) ((Wellauer, Dawid et al. 1976), (Holland, Lowe et al. 2016) (Guetg, Lienemann et al. 2010) (Santoro, Schmitz et al. 2010)).

Early studies using restriction digestion followed by Southern blot hybridization have shown heterogeneous pattern among individuals, suggesting that the distribution of fragments within one size class is characteristic for individuals. Remarkably, identical patterns of fragment size and distribution were also found the same individual (reviewed in (Erickson and Schmickel 1985)), providing a first indication rRNA sequence heterogeneity within the same genome. The restriction fragment length polymorphism is associated with a variable number of enhancer repeats (9, 10, 11, 12 and 22), which seems to be a regulatory sub-domain in the rRNA gene array. Indeed, Santoro et al. showed that the long non-coding RNA IGS-rRNA is transcribed from a specific subclass of rRNA genes containing nine enhancer repeats (Santoro, Schmitz et al. 2010). Since transcription of IGS-RNA and maturation of pRNA is required for NoRC-dependent heterochromatin formation and transcriptional silencing (Savic, Bar et al. 2014), it has been suggested that specific subclasses of rRNA genes are involved in the regulation of rRNA gene silencing and nucleolar organization.

In this work, we aimed to determine whether rRNA gene variants containing different SNPs are distributed differently amongst individuals and if they might have a certain regulatory function, such as the regulation of epigenetic and transcriptional state. For this analysis, it was necessary to establish a quantitative system to measure the SNPs at +44 (rRNA gene variants +44A/T/G) and at -104 (rRNA gene variants -104A/C). We approached this goal by using quantitative PCR (qPCR) and the establishment of primers that specifically amplify rRNA variants containing the corresponding polymorphism (rRNA

gene variants +44A/T/G, -104A/C). Since the amplification efficiency can vary among different primer pairs, it was important to establish a method that also takes into account these possible variations. To do this, I quantified the absolute amounts of amplified rRNA gene variants using as standard curve serial logarithmic dilutions of plasmids containing the 5'-region of the corresponding rRNA gene variants, which contained the specific SNP sequence (+44A/T/G and -104A/C). Since we were unable to clone all the rRNA variants and their combination directly from genomic DNA, we applied site directed mutagenesis on existing plasmid containing rRNA promoter sequences to obtain the complete series of rRNA variants (-104C/+44A (CA), -104C/+44T (CT), -104C/+44G (CG), -104A/+44A (AA), -104A/+44T (AT), -104A/+44G (AG)). The standard curves were obtained through logarithmic dilutions, starting from a plasmid concentration of 100 ng/μl, which was measured by spectrophotometer (Nanodrop). The dilution series comprised 8 dilution steps, ending to a concentration of 0.00001 ng/μl. The initial plasmid DNA concentration was obtained by measuring DNA absorbance. However, this method is quite imprecise for our purpose. Therefore, the amounts of DNA in all the rRNA gene variants standard samples were further assessed by quantitative PCR using primer pairs encompassing a rRNA sequence, which does not contain a polymorphism (-165 forward and +111 reverse, **Figure 8**). In these measurements, we set the values of the concentration of plasmid containing the rRNA gene variant -104C/+44A (CA) as *true* (sample 1, 100 ng/μl) and used the serial dilutions of this DNA as standard curve for all the other samples. As shown in Table 1, the concentration of rRNA gene variants -104A/+44A (AA), (-104A/+44G (AG) and -104A/+44T (AT) were the same as the samples containing rRNA gene variant -104C/+44A (CA). In contrast, the samples containing rRNA gene variants -104C/+44G (CG) and -104C/+44T (CT) had higher concentration (sample 1, 200 ng/ml and 150 ng/ml, respectively). As described below, the results obtained from the comparative quantifications of all rRNA gene standards will allow to quantify the amounts of rRNA gene variants present in cells.

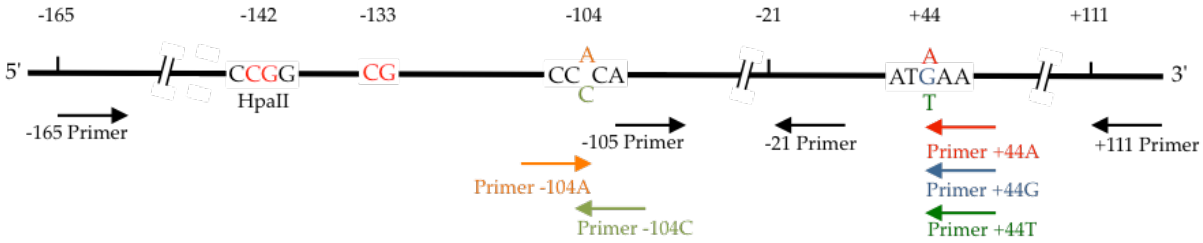

Figure 8 Strategy for the amplification of mouse rRNA gene variants. The respective rRNA polymorphism at position -104 and +44 are marked as well as the CpG sites at position -133 and -142. Arrows represent the primers used for amplification. Black arrows indicate primers that do not encompass sequences with SNPs. Color arrows represent SNP specific primers. At position -142 the analyzed HpaII site is marked.

Name of Standard sample	**Concentration (ng/μl)**	**Name of Standard sample**	**Concentration (ng/μl)**
CA1 **(-104C/+44A)**	100.0	AA1 **(-104A/+44A)**	100.0
CA2	10.0	AA2	10.0
CA3	1.0	AA3	1.0
CA4	0.1	AA4	0.1
CA5	0.01	AA5	0.01
CA6	0.001	AA6	0.001
CA7	0.0001	AA7	0.0001
CA8	0.00001	AA8	0.00001
CG1 (-104C/+44G)	200.0	**AG1 (-104A/+44G)**	100.0
CG2	20.0	AG2	10.0
CG3	2.0	AG3	1.0
CG4	0.2	AG4	0.1
CG5	0.02	AG5	0.01
CG6	0.002	AG6	0.001
CG7	0.0002	AG7	0.0001
CG8	0.00002	AG8	0.00001
AT1 (-104A/+44T)	100.0	CT1 **(-104C/+44T)**	150.0
AT2	10.0	CT2	15.0
AT3	1.0	CT3	1.5
AT4	0.1	CT4	0.15
AT5	0.01	CT5	0.015
AT6	0.001	CT6	0.0015
AT7	0.0001	CT7	0.00015
AT8	0.00001	CT8	0.000015

Table 1 Quantification of rRNA gene variants in standard samples. Plasmids are named according to SNPs at -104 and +44 at rRNA sequences. The first capital letter indicated the nucleotide matching the rRNA gene variant at position -104 whereas the second letter describes the nucleotide matching the rRNA gene variant at position +44. All values were obtained using technical triplicates.

To quantify the different rRNA gene variants, we aimed to generate highly specific primer pairs that amplify exclusively the respective variants. To do so, we reasoned that the complementarity of the last nucleotide at 3' end of each primer should have been key for the specific amplification of each rRNA gene variant since any nucleotide mismatch at primer/DNA template hybrid compromises the polymerase activity of the Taq enzyme. To assess primer specificity, we first tested the respective primer pairs using positive and negative controls (**Figure 9**). As a positive control, I used a plasmid containing the respective polymorphism in the rRNA sequence. As shown in Figure 9A, 0.2 pg of plasmid CA (SNPs at -104C and +44A) was used to test the specificity of the reverse +44A primer, which contains a T nucleotide at the 3' of its sequence. As negative controls, I used the same amount of plasmids CG (SNPs -104C and +44G) and AT (SNPs -104A and +44T), which do not contain the A nucleotide at position +44 in their sequence. The forward primer I used to amplify this sequence encompasses the sequence at position -165 which does not vary amongst rRNA gene variants (**Figure 9A**). The results indicated that the plasmid CA was efficiently amplified (green bar) whereas amplifications of 0.2 pg plasmid containing G and T SNPs at +44 (red bars) were much less efficient (less than 10 fold). Thus, we concluded that the +44A primer specifically amplifies rRNA gene variant +44A. I used the same strategy to assess the specificity of all other primers. As shown in **Figure 9B-C**, the reverse primers +44G and +44T efficiently amplify the corresponding variants (+44G and +44T) but not the other rRNA sequences, indicating specificity in the amplification according to the SNP at +44. I also analysed the specificity of primers for the amplification of rRNA variants containing C or A at -104 site. The amplification with the reverse primer -104C was assessed using the forward primer -165, which as described above encompasses sequence lacking SNPs, and 0.2 pg of plasmid DNA containing -104C rRNA gene variant. As shown in **Figure 9D**, the reverse primer -104C only amplifies the corresponding 104C variant whereas amplification from the same amount of the other rRNA genes variants were not as efficient. Finally, I analysed the specificity of amplification for the forward -104A primer using the reverse primer -21, which encompasses sequence lacking SNPs, and 0.2 pg of plasmid DNA containing -104A rRNA gene variant. Also this primer showed high specificity toward the -104A variant (**Figure 9E**).

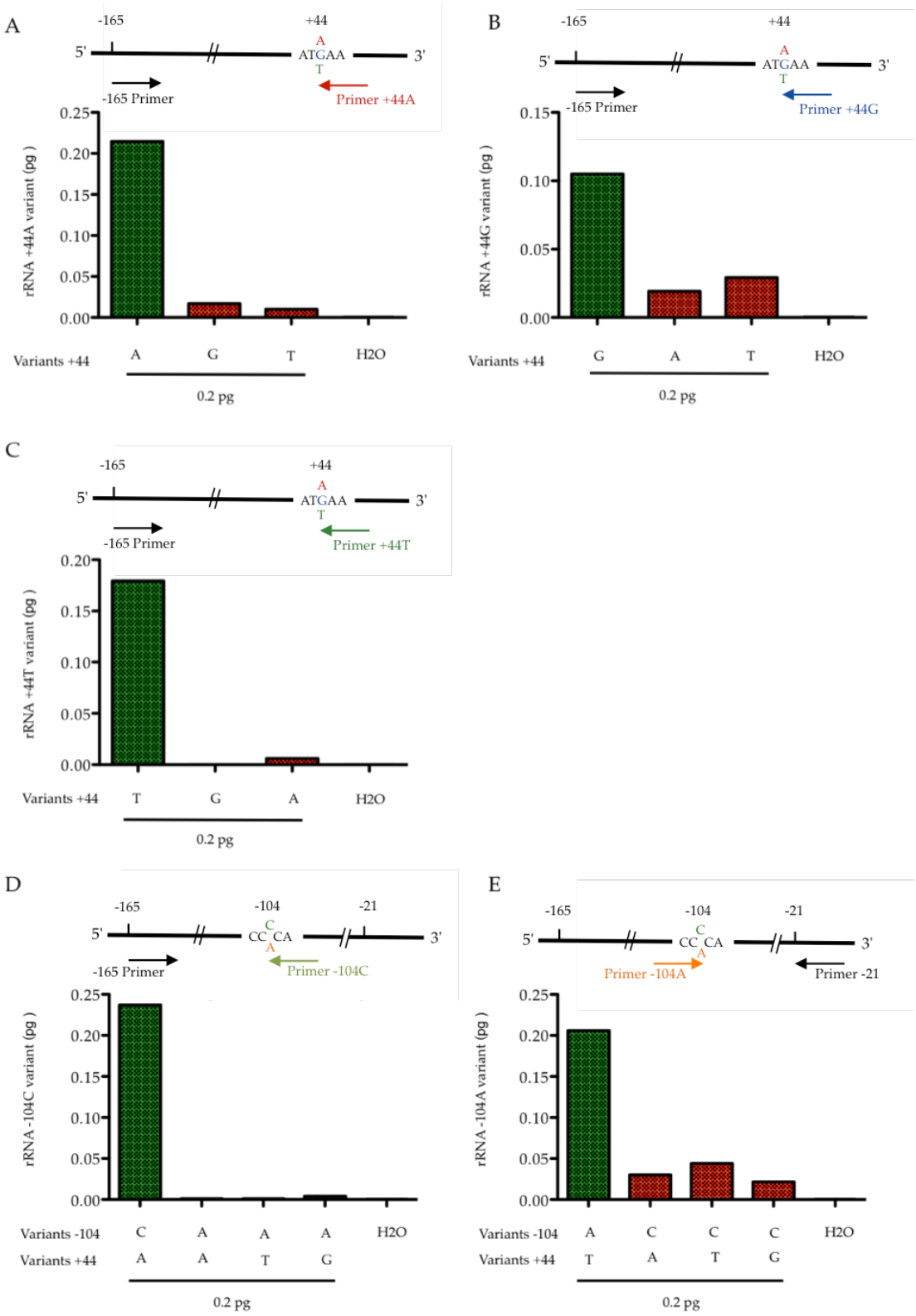

Figure 9 Analysis of primers for the specific amplification of rRNA genes containing single nucleotide polymorphisms at position -104 and +44. (A) Upper panel shows the primers used to amplify the rRNA sequence with SNP +44A. Bottom panel shows qPCR measurements of 0. 2 pg plasmid DNA containing the polymorphisms A (green bar) or G or T (red bars). Values were obtained using the standard curve with rRNA gene variant +44A. Each samples was measured in triplicate. Amplification with reverse primers +44G **(B)**, +44T **(C)**, -104C **(D)** and forward primer -104A **(E)** are shown. Green bars represent values of measurements for the chosen SNP whereas red bars represent values of amplifications of sequences that do not correspond to the SNP.

Unfortunately, the forward primer -104C and the reverse primer -104A did not pass the criteria we set for specific SNP amplification. As shown in **Figure 10A**, the reverse primer -104C can amplify equally well plasmid CA (rRNA gene variant -104C and +44A) and three other plasmids containing the -104A SNP. Similarly, the reverse -104A primer can amplify equally well the plasmid AT (rRNA gene variant -104A and +44T) and three other plasmids containing the -104C SNP. Since these primers did not meet the criteria for primer SNP specificity they were excluded from all further analyses.

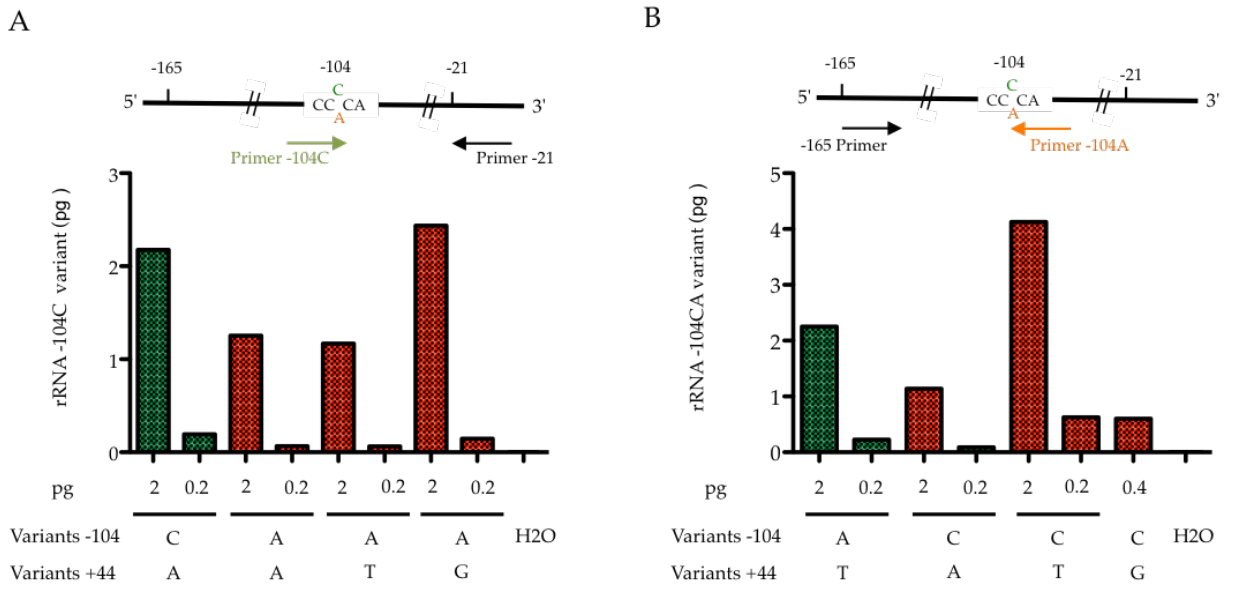

Figure 10 Analysis of primer pairs for single nucleotide polymorphisms at position -104. (A) Upper panel shows the primers used to amplify the rRNA sequence with SNP -104C. Bottom panel shows qPCR measurement of 2 and 0.2 pg plasmid DNA containing the polymorphisms -104C (green bars) or -104A (red bars). Values were obtained using the standard curve with rRNA gene variant -104C, +44A (CA). Each sample was measured in triplicate. **(B)** Upper panel depicts the primers used to amplify the rRNA sequence with SNP -104A. Bottom panel shows qPCR measurement of 2 and 0.2 pg plasmid DNA containing the polymorphisms -104A (green bars) or -104C (red bars). Values were obtained using the standard curve with rRNA gene variant -104A, +44T (AT). Each samples was measured in triplicate.

6.2 Distribution of single nucleotide polymorphisms in different cell lines

The establishment of a SNP specific qPCR represents an important tool to identify and quantify the different rRNA gene variants and to determine the rRNA gene heterogeneity between different individuals and within the same individuals. With this aim, I analyzed the following cell lines and tissues: ESCs cultured in 2i/LIF (ESCs+2i, a cell line currently in use in the Santoro laboratory), ESCs cultured in serum/LIF (ESCs+serum, a kind gift of Constance Ciaudo, ETH Zurich), mouse embryonic fibroblast cells (NIH3T3) and tissues obtained from phalangeal biopsies of C57BL/6 mice (C57BL/6-tissue). It is to note that the two embryonic stem cell lines share the same genetic background since they both derive from the *mus musculus* strain 129/Ola.

I isolated genomic DNA and analyzed the amounts of rRNA gene variants using the SNP specific primers described in section **6.1** (**Figure 8**). To ensure the specificity of the primers in recognizing the correct variants, all the amplifications have been performed with plasmids containing the corresponding rRNA gene variant (positive control) and rRNA gene variants with other SNPs at the same position (negative control). Since the amplification efficiency of different primer pairs can vary, we calculated the values according to the standard curves, which represent serial dilutions of different plasmid DNA (**Table 1**). Using the standard curves described in section **6.1** the obtained values enabled us to analyze absolute amounts of rRNA variant/sample, the sum of which represents the total amount of rRNA genes. To increase the accuracy of the results, all the samples were measured in triplicates.

The data obtained from these measurements revealed that the rRNA variants are indeed differently distributed among different cell lines. Analyses of SNP at +44 site (**Figure 11A**) indicate that in ESCs+2i the +44A-rRNA variant is the most abundant one (49.5%) whereas +44T-rRNA sequence is the least frequent variant (13.9%). The amount of +44G-rRNA variant (36.4%) is comparable to the amount measured in ESCs+serum (35.1%). Interestingly, ESCs+serum lack the +44T-rRNA variant. The absence of +44T-rRNA variants in ESCs+serum line is an unexpected result since both ESC lines derive from the same mouse strain and hence must have the same genotype. As discussed in the Discussion chapter, we reason that during culturing the ESCs+serum line underwent genome instability such as the loss of one or more chromosomes or deletion of defined

regions. A karyotype analyses of these cells is currently on going. The In C57BL/6-tissue all three variants containing the +44-SNP are present in comparable amounts (44A-rRNA: 35.5%, +44G-rRNA: 28.6%, +44T-rRNA: 36%). In contrast, in NIH3T3 cells, the +44A-rRNA variant is less abundant (21.9%) than +44G-rRNA (38%) and +44T-rRNA (40%), which are almost equally represented (**Figure 11A**).

Quantification of rRNA gene variants containing SNPs -104C and -104A revealed unexpected results (**Figure 11B**). Both ESCs+2i and ESCs+serum lack the -104A-rRNA, suggesting the existence of only a limited combination of -104 and +44 SNPs among the rRNA genes. We therefore conclude that ESCs+2i contain three types of -104/+44 variants (-104C/+44A (CA), -104C/+44T (CT) and -104C/+44G (CG) whereas ESCs+serum contain only two -104/+44 rRNA variants: -104C/+44A (CA) and -104C/+44G (CG)). In contrast, we found the presence of both SNPs A and C at -104 in C57BL/6-tissue and NIH3T3 cells. The -104A-rRNA variant is less abundant (C57BL/6: 38.2%, NIH3T3: 34.6%) than the -104C rRNA variant (C57BL/6: 61.8%, NIH3T3: 65.4%) (**Figure 11B**). Thus, in C57BL/6-tissue and NIH3T3 cells rRNA genes are possibly present in six different combinations of -104 and +44 SNPs (rRNA variants -104C/+44A (CA), -104C/+44T (CT), -104C/+44G (CG), -104A/+44A (AA), -104A/+44T (AT), -104A/+44G (AG).

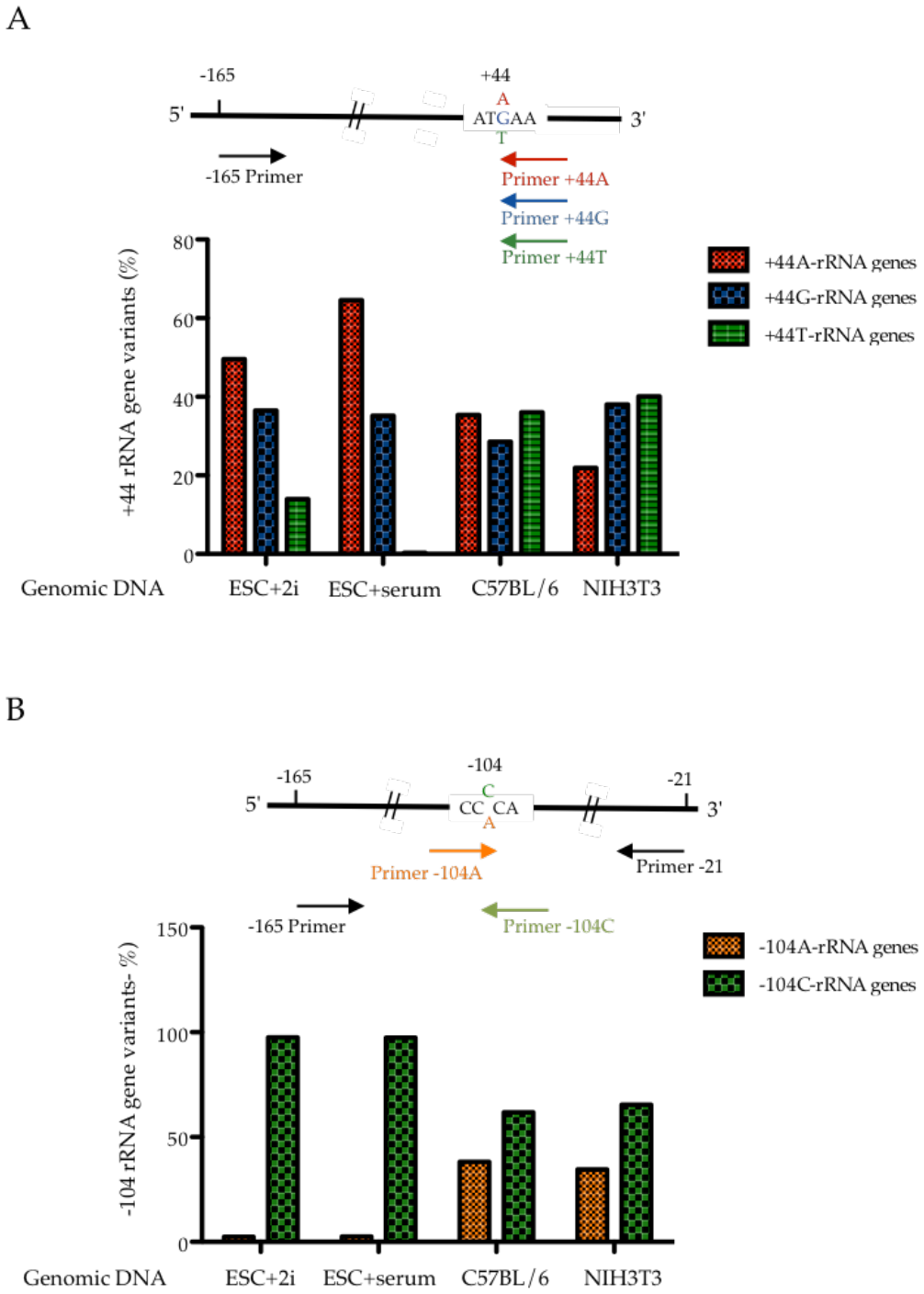

Figure 11 Quantification of -104 and +44 SNP distribution in cell lines and tissues. (A) Upper panel shows the primers used to amplify rRNA sequences with SNP +44. Bar chart shows the abundance of +44A, +44G and +44T rRNA variants. Values represent relative amounts of the respective variants normalized to the total amount of rRNA genes calculated as the sum of the +44 variants amounts. **(B)** Upper panel depicts the primers used to amplify the rRNA sequence with SNP -104. Bar chart shows the abundance of -104A and -104C rRNA variants. Values represent relative amounts of the respective variants normalized to the total amount of rRNA genes as indicated in **A.**

6.3 Methylation of rRNA variants upon embryonic stem cell differentiation

As described in **3.3.3**, upon differentiation of embryonic stem cells (ESCs) a fraction of rRNA genes undergoes silencing through epigenetic modifications mediated by the nucleolar repressor TIP5, a component of NoRC complex (Savic, Bar et al. 2014). In differentiated cells TIP5 is targeted to rRNA gene promoter and recruits DNA methyltransferases with consequent establishment of *de novo* methylation at rRNA genes ((Savic, Bar et al. 2014),(Leone, Bar et al. 2017)). Formation of heterochromatin at a fraction of rRNA genes is critical since abrogation of this process impairs ESC differentiation (Leone, Bar et al. 2017). Although the mechanisms of how this reaction is achieved are quite well understood, it still remains to clarify whether epigenetic silencing at rRNA genes is a random event (i.e. any rRNA gene can be epigenetically silenced) or whether a defined set of rRNA genes is specifically CpG methylated and transcriptionally repressed.

The results described in the previous sections determined that rRNA genes do not share the exact same sequences. We reasoned that the identification and quantifications of SNPs at +44 (+44A/T/G rRNA gene variants) and at -104 (-104A/C rRNA gene variants) will allow us to learn whether a defined class of rRNA genes is more prone to acquire DNA methylation upon ESC differentiation. To measure DNA methylation state at specific rRNA gene variants, we applied a strategy based on the measurement of resistance to the digestion of the restriction enzyme HpaII (**Figure 12A**). The sequence CCGG is recognized by HpaII. However, HpaII digestion is blocked by CpG methylation. Upon digestion of genomic DNA with HpaII, the use of primers encompassing the CCGG site allows the amplification and subsequent quantification of HpaII resistant and therefore methylated sequences. Amplification of sequences lacking CpGs allows the measurement of total DNA, which is then used to calculate the proportion of methylated sequences. The mouse rRNA gene promoter contains two CpGs sequence at -142 and -133 which are both methylated at silent rRNA genes (Santoro and Grummt 2001). CpG dinucleotide at -142 is within the sequence CCGG, which can be digested by HpaII (**Figure 12B**). Therefore, resistance to HpaII digestion at the rRNA promoter sequences is indicative of CpG methylation and this can be measured by quantitative PCR using appropriate primers (Santoro, Li et al. 2002). The proportion of methylated rRNA sequences can be

then measured through normalization of total rRNA sequences obtained with the amplification with primers amplifying the rRNA region from -105 to -1, which lacks CpG residues. To assess equal digestion between samples, we included a further control based on the digestion of pBluescript plasmid DNA, which does not contain methylated CpGs. In each sample genomic DNA and pBluescript were mixed together and incubated with HpaII. The digestion efficiency of pBluescript DNA was calculated through amplification with a primer pair encompassing CCGG sequences (from 2409 to 2587) and normalized to regions lacking HpaII sites (from 2409 to 2544, **Figure 12C**).

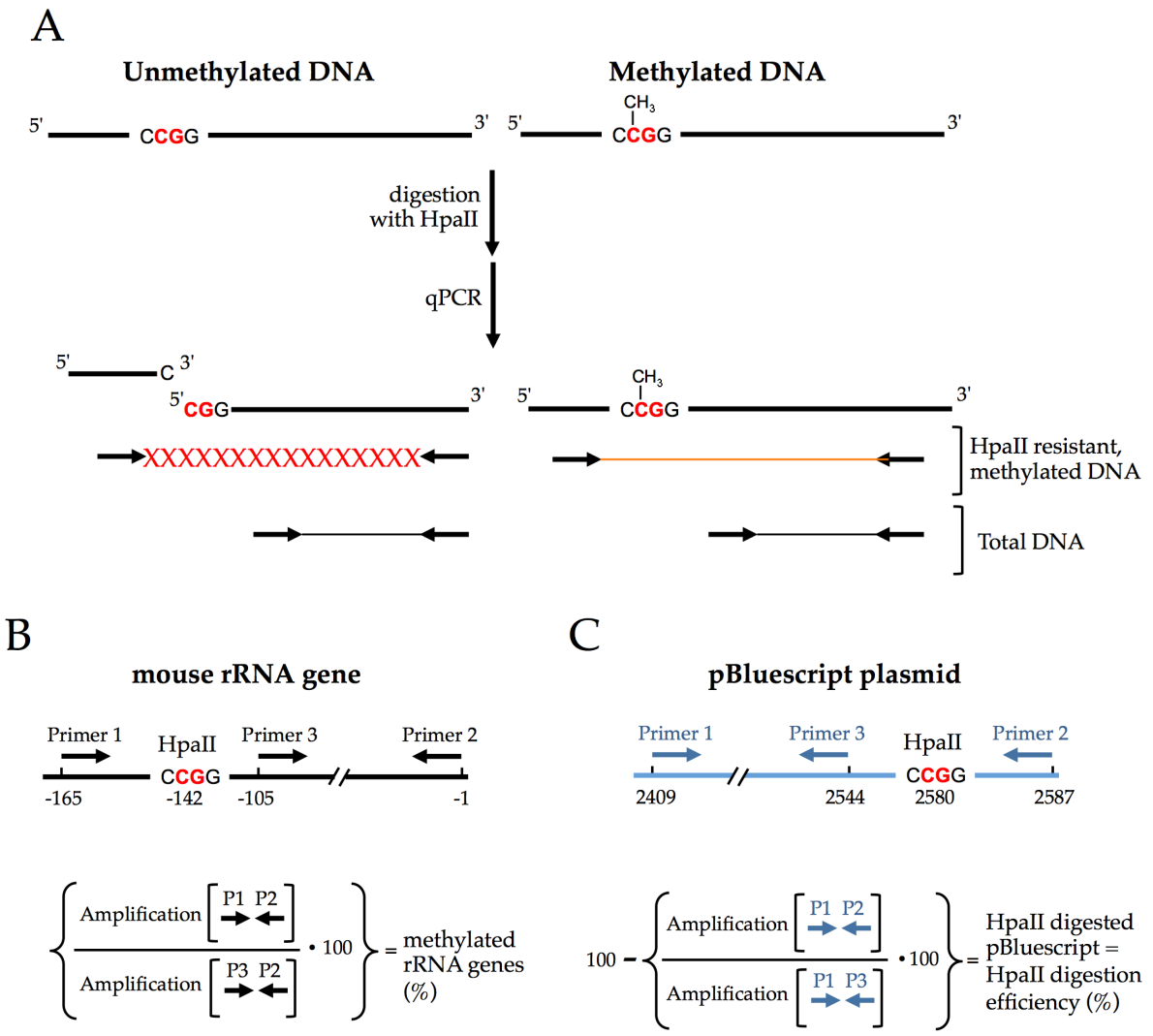

Figure 12. Strategy to measure methylation of rRNA genes. (A) Schema depicts the rationale for the measurement of DNA methylation via the assessment of resistance to the digestion of the restriction enzyme HpaII by quantitative PCR. **(B)** Schema of the rRNA promoter. The HpaII digestion site CCGG is marked at position -142. Black arrows represent the primer pairs used for the amplification. Formula shows how the amounts of methylated total rRNA promoter sequences are calculated (%). **(C)** Schema of the sequence of pBluescript plasmid. The HpaII digestion site is marked at position 2580. Blue arrows represent the primer pairs used for amplification. Formula shows how the amounts of undigested pBluescript sequences, which represents HpaII digestion efficiency, are calculated (%).

We initially measured the amount of CpG methylation at total rRNA genes in ESCs and upon differentiation of ESCs into neural progenitors (NPCs). To generate NPCs we used an established protocol (Bibel, Richter et al. 2004) where ESCs growing in 2i/LIF (ESCs+2i) or serum/LIF (ESCs+serum) were cultured with a neural differentiation media on a not adherent plate. After 4 days culture, cells were grown in the presence of retinoic acid (RA) for further 4 days. As shown in **Figure 13**, NPCs form cellular aggregates (embryonic bodies) that grow in suspension.

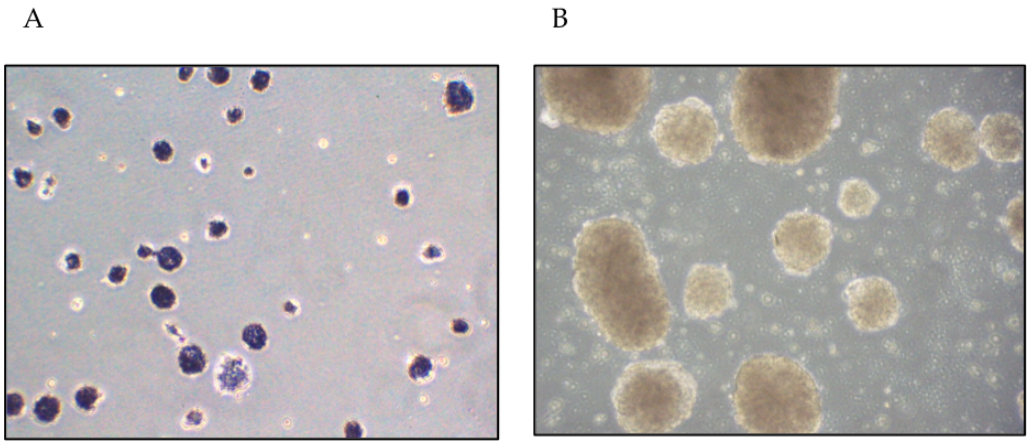

Figure 13 Representative pictures of (A) embryonic stem cells (ESCs) stained for alkaline phosphatase (AP) and (B) neural progenitor cells (NPCs).

To measure DNA methylation at rRNA genes, we performed HpaII digestion of purified genomic DNA from ESCs+2i, ESCs+serum, and the corresponding derived NPCs (NPCs/2i and NPCs/serum). To assess an equal efficiency of HpaII digestion among samples, each genomic DNA was digested in the presence of unmethylated pBluescript plasmid. Efficiency of HpaII digestion was measured by calculating the amounts of undigested pBluescript sequences using primers encompassing the HpaII site at position 2580 (**Figure 12B**). As shown in **Figure 14A**, our experimental conditions allowed a high efficient HpaII digestion, which was very similar among all treated samples (ESCs+2i: 99.6%, NPCs/2i: 99.9%, ESCs+serum: 99.7% and NPCs/serum: 99.7%).

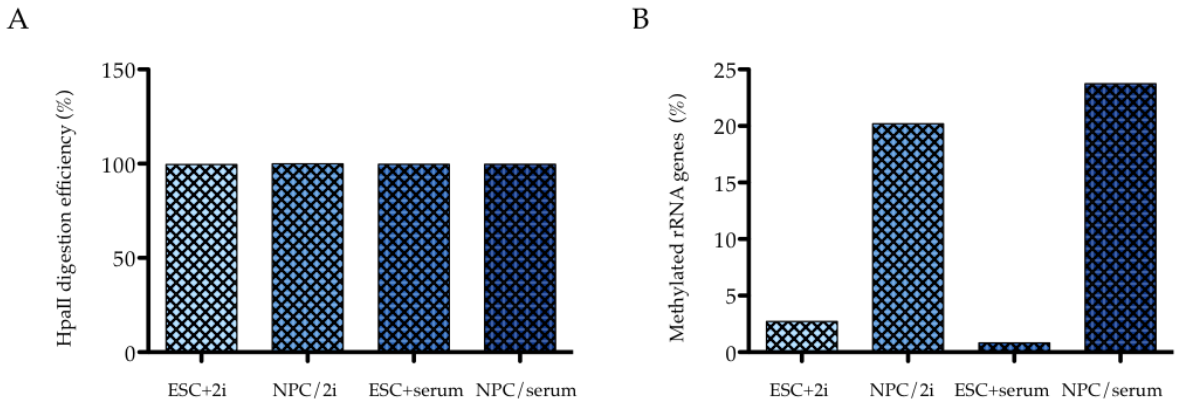

Figure14 rRNA genes undergo CpG methylation upon ESC differentiation into NPCs. (A) Efficiency of HpaII digestion measured by amplifications of pBluescript sequences as indicated in Figure 12C. **(B)** Methylation of rRNA genes in ESCs+2i, NPCs/2i, ESCs+serum, NPCs/serum obtained through the amplification of genomic DNA digested with HpaII. Values were obtained by normalizing the amounts of methylated, HpaII resistant rRNA genes sequences to total rRNA genes as shown is Figure 12B. Measurements of amplified sequences were calculated using a series of logarithmic dilutions of pBluescript plasmid and plasmids containing RNA gene sequences (see Table 1).

As described in **3.2**, the genome of the inner cell mass (ICM) cells contains low levels of DNA methylation (Smith, Chan et al. 2012). Similarly, derivation of ESCs from ICM through culturing with 2i allows the establishment of a *in vitro* ground state which coincides with a hypomethylated genome. In contrast, ESCs cultured in serum showed high CpG methylation levels (Marks and Stunnenberg 2014). We initially analysed the CpG methylation content of total rRNA genes in ESCs and NPCs (**Figure 14**). We found that only few rRNA genes (2.72%) in ESCs+2i are methylated, which is consistent with recent results showing that the low methylation content in ESCs+2i is due to the down regulation of the *de novo* methyltransferases DNMT3b and DMNT3a ((Leitch, McEwen et al. 2013),(Ficz, Hore et al. 2013)) (**Figure 14B**). Remarkably, also in the case of ESCs+serum, which contain elevated CpG methylation due to elevated Dnmt3a and Dnmt3b expression levels, only very few rRNA genes are methylated (0,8%), supporting an important role in the maintenance of rRNA genes in a hypomethylated state that is linked to the pluripotency state as recently suggested by (Leone, Bar et al. 2017). Consistent with previous results (Savic, Bar et al. 2014), upon differentiation into NPCs, a fraction of rRNA genes acquires epigenetic silencing through CpG methylation. As shown in **Figure 14B**, both NPCs derived from ESCs+2i and from ESCs+serum showed *de novo* methylation at a fraction of rRNA genes, 20.2% and 23.7% respectively. These values are very similar to the levels of rRNA gene methylation recently reported in mouse brain

tissues (30%) (Savic, Bar et al. 2014), underscoring a functional link between the establishment of CpG methylation at rRNA genes and early development.

The results shown in the previous sections determined the existence of certain rRNA gene subclasses, which do not share the same sequences (rRNA gene variants +44A/T/G and rRNA gene variants -104A/C). These rRNA gene variants showed a distinct distribution pattern in different cell lines (**Figure 11**). To determine whether a defined set of rRNA genes is specifically CpG methylated and transcriptionally repressed upon ESC differentiation, we measured DNA methylation levels of the different rRNA gene variants in ESCs and NPCs, applying the experimental strategy shown in **Figure 15A**. Because of the lack of a SNP at -104 in both analyzed ESC lines, all measurements were performed on +44 rRNA gene variants. Genomic DNA from the two ESC lines (ESCs+2i and ESCS+serum) and the derived neural progenitors (NPCs/2i and NPCs/serum) were digested with HpaII and the amounts of CpG methylated +44 A/T/G-rRNA variants were obtained by quantitative amplifications using the forward -165 primer and the reverse +44 SNP reverse primers. These values were then normalized to the total amount of rRNA quantified through amplifications of rRNA sequence lacking SNPs using forward -105 and the -21 reverse primers (**Figure 15**).

In these experiments, we detected 5.7% methylation in ESCs+2i, which was mostly present at +44G variant whereas +44A and +44T variants were not methylated (**Figure 15C**). Interestingly, although the +44A variant is the mostly abundant in this ESC line (49.5% of total rRNA genes, **Figure 15B**), upon differentiation into NPCs it acquires less *de novo* methylation (9.1%) compared to the other variants (+44G-rRNA: 16.3% and +44T-rRNA: 17.9%) (**Figure 15B,C**). On the other hand, the +44T-rRNA variant, which is the least abundant variant (14%) in ESCs+2i, seems to be more prone to acquire CpG methylation upon differentiation (17.9%) than the other rRNA gene variants. As described in section **6.2**, the cell line ESC+serum lacks the +44T-variant (**Figure 15D**). Analysis in ESCs+serum revealed very low and comparable levels of DNA methylation at +44A (1%) and +44G-rRNAs (1.3%) (**Figure 15E**). Upon differentiation into NPCs, although the +44A variant is almost 2-fold more abundant than +44G variant (64.5% of total rRNA genes), it contributes to DNA methylation only for 9% whereas methylated +44G-rRNAs represent 15.9% of methylated sequences. Taken together these results indicated that +44A-rRNA genes are less prone to be methylated whereas +44G and +44T-variants are the major

targets for *de novo* methylation. These data strongly indicate that upon ESC differentiation methylation at rRNA genes is not randomly distributed but some rRNA gene variants are more prone to acquire CpG methylation.

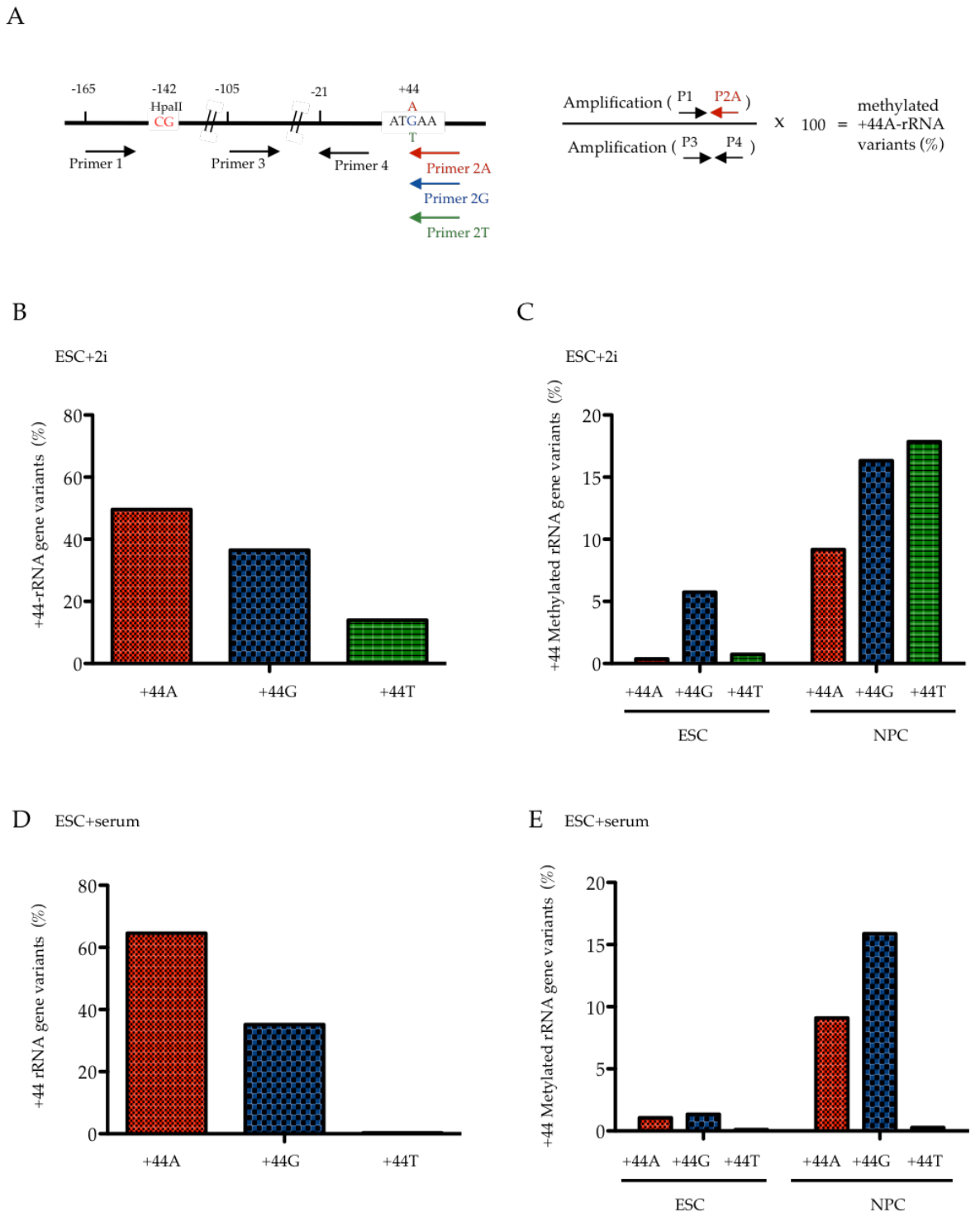

Figure 15 Distribution of CpG methylation at rRNA variants during ESC differentiation. (A) Left panel depicts primers used to amplify +44-rRNA variants and the strategy to measure methylation at -142. Right panel shows the example of how methylation levels at +44A-rRNA genes were calculated. In this case, the amounts of +44A-rRNA sequences resistant to HpaII digestion (i.e. methylated at CpG -142) obtained through amplifications with primers 1 (P1) and +44A primer 2 (P2A) were normalized to total amount of rRNA genes using primer 3 (P3) and primer 4 (P4) **(B)** Bar chart shows the abundance of +44A, +44G and +44T rRNA variants in ESCs+2i. Values represent relative amounts of the respective variants normalized to the total amount of rRNA genes calculated as the sum of the +44 variants amounts. **(C)** Methylation levels at rRNA variants in ESCs+2i and in cells after 8 days differentiation toward neuronal lineage (NPCs, neuronal progenitors). Bar chart represents the amounts of rRNA gene variants resistant to HpaII digestion (i.e. methylated at CpG -142). Measurements were performed in triplicates and values were calculated as described in **(A)**. **(D)** Bar chart shows the abundance of +44A, +44G and +44T rRNA variants in ESCs+serum line. **(E)** rRNA variant methylation in ESCs+serum and NPCs derived upon 8 days of differentiation

6.4 Expression of rRNA gene variants

DNA methylation is an epigenetic mark associated with gene silencing and repressed transcription (see chapter **3.2**). Cytosine methylation can change the DNA structure and alter the binding of DNA binding proteins, which may modulate transcription. Some DNA binding proteins, as for example certain transcription factors are only able to interact with their target sequence if not methylated. Hence CpG methylation directly abolishes their interaction with the DNA and leads to lower transcription levels (Clark, Harrison et al. 1997). rRNA genes represent an important example of how DNA methylation affects gene transcription. Methylation of the rRNA gene promoter affects chromatin structures and impairs the association of the RNA Polymerase I (Pol I) transcription factor upstream bing factor (UBF), thereby preventing assembly of pre-initiation complex and repressing transcription (Santoro and Grummt 2001).

The data described so far have shown that CpG methylation of rRNA genes in ESCs+2i and ESCs+serum is very low and rRNA genes acquired methylation only upon differentiation (**Figures 14,15**). Moreover, the results also showed that methylation at rRNA genes is not randomly distributed but +44T and +44G-rRNA variants are more prone than +44A-rRNA variants to acquire CpG methylation. To understand whether the acquisition of DNA methylation depends on the transcription state and whether increased CpG methylation during differentiation affects transcription we measured rRNA expression levels in ESCs and NPCs.

To measure the rRNA gene transcription, we purified total RNA from ESCs+2i and NPCs/2i. After reverse transcription (RT) with random examer primers, the cDNA was used for qPCR in order to quantify the expression levels of rRNA genes. The 45S pre-rRNA is the transcript originating from rRNA genes (**Figure 16A**). The half-life of 45S pre-rRNA is short since upon transcription this rRNA is immediately processed to generate 18S and 28S rRNA, which in turn have ca. 2-days half-life (Moss, Langlois et al. 2007)). Therefore, the quantification of rRNA transcription is generally analyzed through measurements of 45S pre-rRNA levels obtained by the amplification of sequences encompassing the 1st processing site, located between +500 and +700 relative to transcription start site (**Figure 16A**) (Santoro 2005). Since the aim of this study was also to measure transcription of +44 rRNA variants by quantitative PCR, we had no alternative

than to measure rRNA transcripts encompassing the 5'-rRNA sequences (from +1 to +65) (**Figure 16A**). Nevertheless, it is very unluckily that these transcripts have a long-half life since they are not part of the ribosomes. To ensure that quantifications of rRNA sequences at the 5' rRNA regions correspond to the values obtained by quantifying the sequence encompassing the 1st rRNA processing site (45S-pre-rRNA), we measured rRNA transcripts in ESCs and NPCs and compared the values obtained with the amplification using the two respective primer pairs (**Figure 16B,C**). We did not observe remarkable difference between 45S-pre-rRNA and +1/+66 rRNA levels in ESCs and NPCs, suggesting that values obtained with both primer pairs well represent rRNA transcription measurements.

The data described in the previous sections determined that rRNA genes in ESCs+2i contain low methylation (2.72%), whereas upon differentiation 20.2% of the rRNA genes acquire *de novo* methylation (**Figure 14 and 15**). Although it would be expected a decrease of rRNA transcription in NPCs, in this experiment I did not observe any change. The reason of this result is not clear since previous data have shown downregulation of rRNA transcription upon differentiation (Savic, Bar et al. 2014).

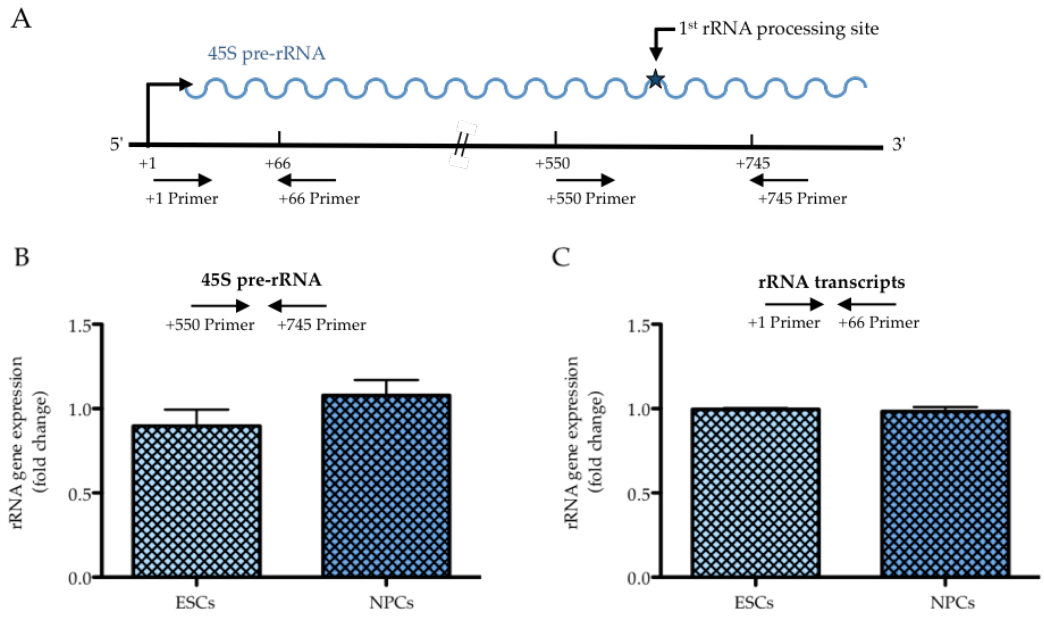

Figure 16 Measurements of rRNA transcription in ESCs and NPCs. (A) Schema represents the 5' region of one rRNA gene repeat, the 45S pre-rRNA (blue line) and the 1st rRNA processing site (blue star). Primer and their position are represented as black arrows. The blue arrow shows the transcription start site. **(B)** rRNA gene expression measured with the primer pair (+550 / +745) amplifying the region encompassing the 1st rRNA processing site, which represent the 45S pre-rRNA. Values are normalized to *rps12* mRNA levels. **(C)** rRNA gene expression measured with the primer pair (+1/ +66) amplifying 5'-region of rRNA transcript. Values are normalized to *rps12* mRNA levels

According to the data shown in **Figure 15B**, in ESCs+2i the +44G-rRNA variant is the most frequently methylated gene variant (5.7%), whereas +44A and +44T-rRNA variants are mostly unmethylated. Upon differentiation, *de novo* CpG methylation occurs particularly at the +44G-rRNA- and +44T-rRNA variants (16.3% and 17.9% respectively), while only 9.1% of the rRNA-A gene variant acquires methylation. Since CpG methylation during differentiation is not equally distributed amongst the different rRNA gene variants, we reasoned to measure expression levels of rRNA subclasses containing a SNP at +44 to determine whether CpG methylation in differentiated cells affects their transcription. Analysis of +44-rRNA variants indicated that in both ESCs and NPCs transcripts originated from +44G-rRNA genes are about the half of +44A or +44T-rRNAs (**Figure 17**). Considering the relative amounts of +44-rRNA variants and hypothesizing that all rRNA genes transcribe in ESCs because of the lack of the repressive DNA methylation, +44T-rRNA genes are the most active genes in both ESCs and NPCs (4-fold higher than +44A and 7-fold higher that +44G variants). Accordingly, +44G-rRNA variants are the less transcribed genes (ca. 50% less active than +44A-rRNA variants). These results indicate that the degree of methylation does not correlate with transcription levels since the variants with higher methylation (+44G and +44T) have a similar down-regulation as the +44A-rRNA variants, which are less methylated upon differentiation. Moreover, they indicate that the genes with the highest transcription in ESCs (+44T-rRNA variants) undergo preferential de novo methylation upon exit from pluripotency. Although these descriptive data still require further validation using different experimental settings (i.e. time-point analysis during differentiation), they indicate that sequence heterogeneity at rRNA genes might also affect the transcriptional efficiency, a working model that will be object of future studies in our lab.

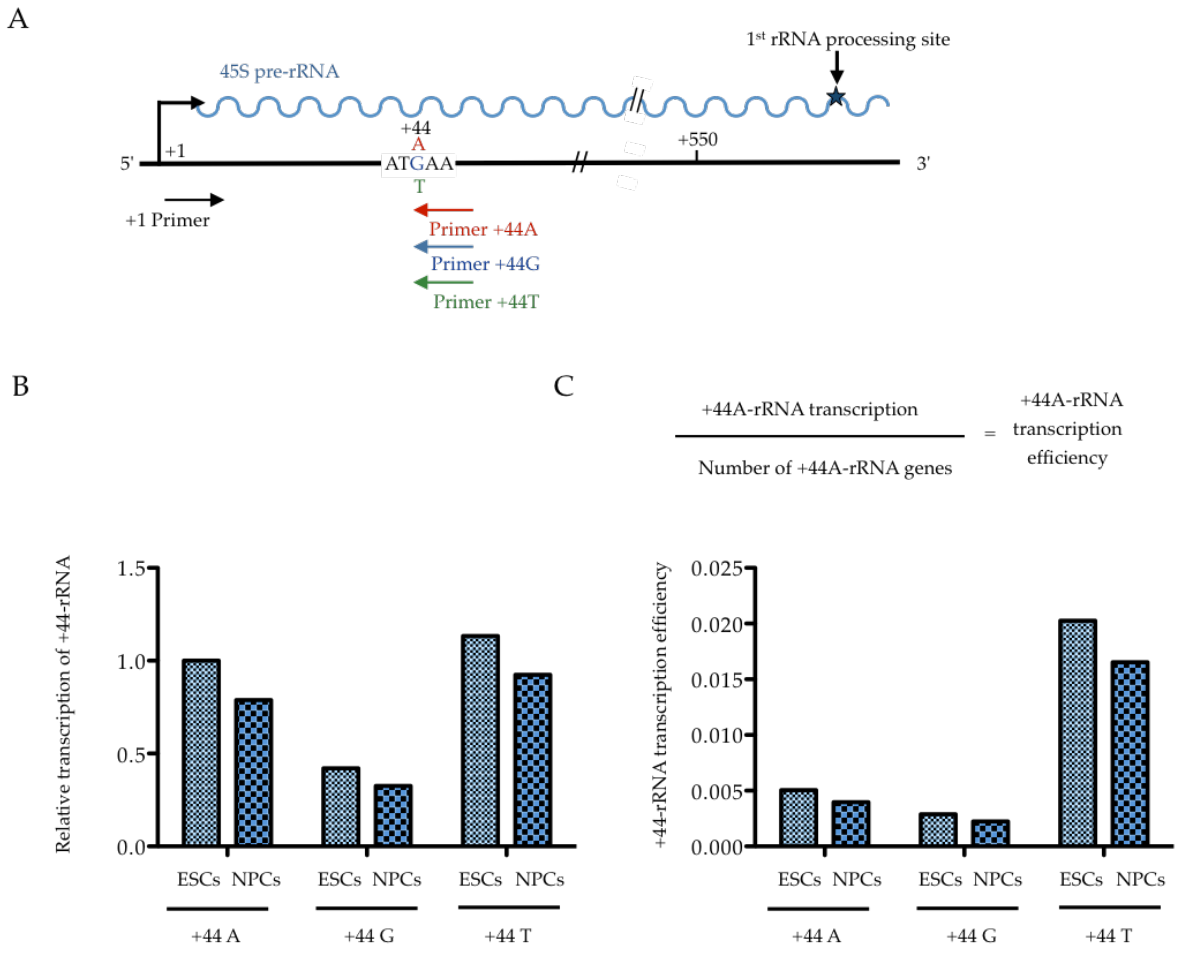

Figure 17 Analysis of the rRNA gene variant expression in ESCs and NPCs. (A) Schema represents the 5' region of one rRNA gene repeat, the 45S pre-rRNA (blue line) and the 1st rRNA processing site (blue star). +44-SNP primers used for RT-qPCR are represented as arrows (orange, +44A; blue, +44G; green, +44T). **(B)** Relative transcription of +44-rRNA variants. Data were normalize to the amounts of the +44A-rRNA transcript in ESCs. **(C)** Transcription efficiency of +44-rRNA variants. Values were calculated by normalizing the amounts of transcripts to the number of rRNA variant. The formula shows the example how +44A-rRNA transcription efficiency has been calculated.

6.5 CpG methylation of rRNA gene variants in NIH3T3 cells

The results described in **Figure 11** determined that the distribution of rRNA gene variants vary between individuals. For example, whereas in ESCs+2i +44A-rRNA gene variant is the most abundant variant (49.5% of all rRNA genes), in NIH3T3 cells it is the least frequent variant (21.9%). In contrast, +44T-rRNA variant is the most abundant in NIH3T3 (40%), whereas the ESCs+serum line lacks this variant. Moreover, the results in **Figure 15** revealed that in ESCs+2i and ESCs+serum defined classes of rRNA genes (+44G and and +44T-rRNA variants) are more prone to acquire DNA methylation upon ESCs differentiation than the +44A-rRNA genes. Finally, whereas both analyzed ESCs line lack

a polymorphism at -104, rRNA genes in NIH3T3 cells (also in mouse C57BL/6 genome) contain either A or C at -104 (**Figure 11B**).
To determine whether also in NIH3T3 cells certain rRNA gene subclasses are more likely to acquire CpG methylation, we measured the resistance to HpaII digestion at the promoter of the different rRNA variants. We initially quantified the amount of CpG methylation at total rRNA genes in NIH3T3 cells and found that 32.6% of rRNA genes are methylated, which is very close to previous results showing ca. 40% rRNA gene methylation in NIH3T3 cells (Santoro and Grummt 2001) (**Figure 18B**).

To determine whether a certain subclass of rRNA genes shows higher levels of CpG methylation respect to others, we have first measured methylation levels of the different rRNA gene variants containing the SNP at position +44. Our results described in section **6.3** (**Figure 15**) have shown that upon differentiation of ESCs+2i and ESCs+serum the +44A-rRNA gene variants acquire less CpG methylation compared to +44G and +44T variants. Interestingly, +44A rRNA variants are also less methylated (2% off all rRNA genes) than +44G-rRNA genes (8.3%) and +44T-rRNA genes (18.7%). The fact that a large fraction of methylated sequences are +44T-rRNA genes in NIH3T3 cells is consistent with our data showing that upon differentiation of ESCs+2i (which contain +44T-rRNA genes) this variant has the highest level of methylation (17.9%) (**Figure 18D**). These data further support a model in which certain subclasses of rRNA genes is more prone to acquire DNA methylation than others.

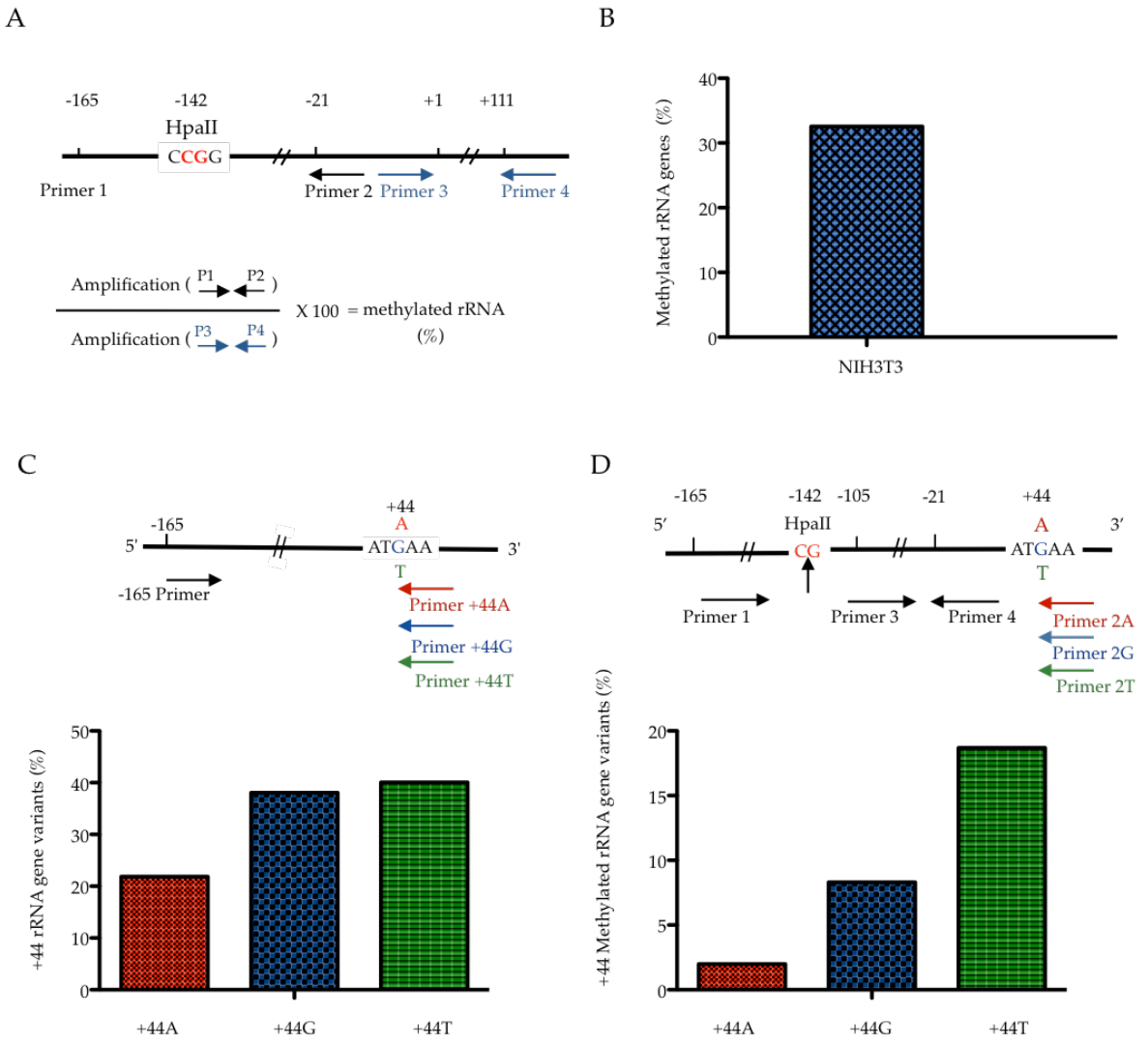

Figure 18 CpG methylation levels of rRNA genes in NIH3T3 cells. (A) Shows the strategy used to measure CpG methylation of total rRNA genes in NIH3T3 cells. **(B)** Shows values of CpG methylation at rRNa genes. **(C)** Quantification of +44-rRNA variants in NIH3T3 cells. Upper panel shows the strategy used to measure the specific variants. Measurements were normalized to total amount of rRNA genes calculated as the sum of the amplified variants. **(D)** Methylation of +44A-rRNA gene variants in NIH3T3 cells. Upper panel shows the strategy used to measure methylation. Measurements were normalized to total amount of rRNA genes calculated as shown in panel **A**

As described above, in contrast to ESCs+2i and ESCs+serum lines, rRNA genes in NIH3T3 contain a further SNP at -104 (A or C). Remarkably, recent data have shown the -104A-rRNA variant undergoes methylation in the offspring of mice subject to protein-restricted diet during pregnancy and that this specific methylation correlates negatively with weaning weight (Holland, Lowe et al. 2016).

To measure CpG methylation, we had to adapt our strategy since the forward -104C and the reverse -104A primers did not pass our quality control, being unable to amplify the corresponding variants in a specific manner (**Figure 10**). These primers would have allowed to use a strategy similar to the one used for +44 variants. This is particularly evident for the measurement of methylation at -142 CpG in the -104A variant, which could

have been measured through amplification of HpaII digested DNA with the forward -165 (-165 for) and the reverse -104A primers. Therefore, the lack of specificity of the reverse -104A primers does not allow a direct measurement of methylation of -104A-rRNA variant. On the other hand, the reverse -104C (-104C rev), which passed all the criteria for the specific SNP amplification, could be used in combination with the -165 for primer to amplify HpaII resistant (i.e. methylated at -142 CpG) -104C-variants. However, since the forward -104C primer lacked specificity, the total amount of -104C variants from HpaII digested DNA could not be measured in the same sample. We therefore decided to measure methylation of -104C-rRNA gene variants by comparing the amounts of DNA amplified with -165 for and -104C rev primer pair from NIH3T3 genomic DNA digested and not digested with HpaII (**Figure 19A**). Accordingly, values obtained from HpaII digested DNA will refer to methylated rRNA sequences whereas values obtained from untreated DNA represent the total amounts of rRNA genes. To ensure that the amount of genomic DNA is equal in both samples (with and without HpaII), I included an additional control to normalize the total amount of rRNA genes through the amplification of rRNA sequences not containing a HpaII (Used primer pair: +1 forward and +111 reverse). Finally, we reasoned that the amounts of methylated -104A-rRNA variants could be obtained by subtracting the amounts of methylated -104C-variant from the total amount of methylated rRNA genes (**Figure 19A**).

The results shown in **Figure 19B** indicate that in NIH3T3 cells the -104C-rRNA variant is more abundant (65.4%) than the -104A-rRNA genes (34.6%), which is in agreement with the measurements described in **Figure 11**. Methylation analyses showed that the amounts of methylated -104C-variants are very similar to the total amounts of methylated rRNA genes, indicating that -104A variant is not methylated (**Figure 19C**). These results are consistent with our analysis of +44-rRNA gene variants, showing that methylation at rRNA genes is not randomly distributed and further support a model in which certain subclasses of rRNA genes is refractory to CpG methylation.

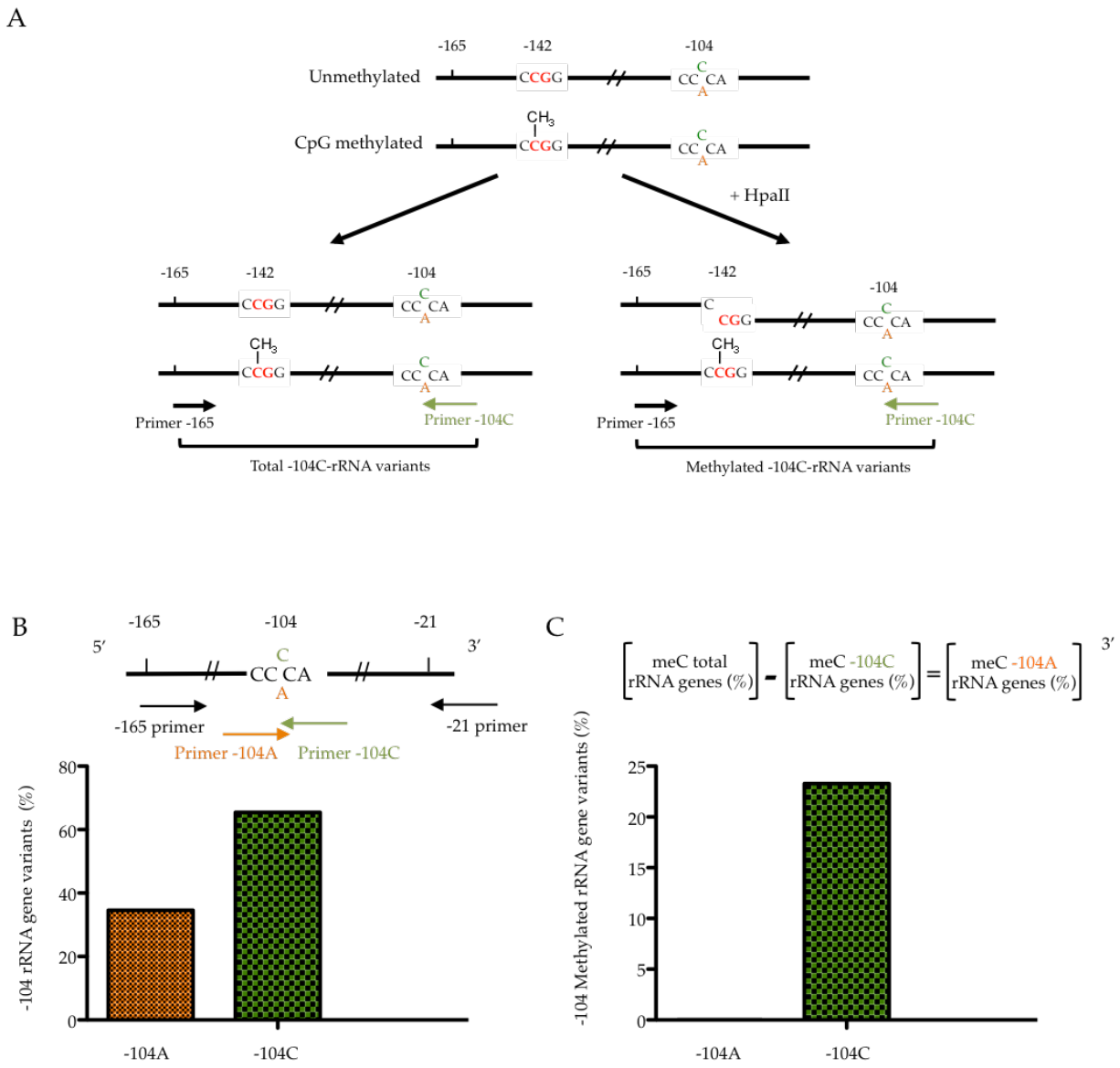

Figure 19 Analysis of methylation of -104-rRNA variants in NIH3T3 cells (A) Experimental strategy to quantify methylation levels of -104C-rRNA variants. HpaII digested and mock control genomic DNA from NIH3T3 cells were amplified with the indicated primers. The values obtained by amplification of mock control represent the total amount of -104C-rRNA variants. The values from HpaII digested DNA represent the methylated -104-genes. **(B)** Quantification of -104-rRNA variants in NIH3T3 cells. Upper panel shows the strategy used to measure the specific variants. Measurements were normalized to total amount of rRNA genes calculated as the sum of the amplified variants. **(C)** Relative methylation levels of -104-rRNA gene variants. Values for methylated -104C-variants were obtained as described in **A.** The upper panel describes how the levels of methylated -104A-rRNA variants were calculated.

7 Discussion

An important concept that is emerging in modern biology is the cellular heterogeneity. This is particularly evident in cancer biology, since heterogeneity poses practical challenges for building accurate clinical models–particularly ones based on population-averaged measurements–to guide diagnosis and treatment of the disease ((Campbell and Polyak 2007), (Altschuler and Wu 2010))

In contrast to cancer cells, which accumulate genetic mutations during disease progression, healthy cells from an individual organism are considered to contain the same genome and hence do not display genetic heterogeneity. However, healthy cells from the same individual organism can have in their genome a certain degree of sequence variations, which can be found between paternal and maternal alleles and in particular at repetitive sequences such as satellite DNA, transposable elements and rRNA genes. Some of these sequence are very abundant such as major satellite repeats, which are represented about 10,000 times in every chromosome and make up ~10% of the mouse genome (Garagna, Zuccotti et al. 2002) or interspersed repeats (LINE, SINE, LTRs and ERV elements), which comprise ~43% of the mouse genome (Bulut-Karslioglu, De La Rosa-Velazquez et al. 2014). Some of the biggest technical challenges that are associated with next-generation sequencing projects are caused by repetitive DNA. From a computational perspective, repeats create ambiguities in alignment and in genome assembly, which, in turn, can produce errors when interpreting results (Treangen and Salzberg 2011). The repetitive nature of rRNA genes is not an exception. So far, only one mouse and one human rRNA repeat has been sequenced (Third Party Annotation accession numbers BK000964 and U13369), and these are generally considered to be the sequences of all rRNA genes (Santoro, Schmitz et al. 2010). However, the fact that all rRNA repeats do not share the same sequence was already evident in early studies showing heterogeneous patterns of restriction rRNA fragments in the non-transcribed spacer DNA among individuals such as BALB/c, C57BL and *Mus poschiavinus* strains and human leukemic blood cells (Arnheim and Southern 1977). Heterogeneity in rRNA sequences was also found in the same individuals (Cory and Adams 1977) (Erickson and Schmickel 1985), indicating that one genome can contain several rRNA gene subclasses

(rRNA variants). However, whether sequence variation in rRNA genes affects gene regulation remains yet elusive.

The aim of this work was to create tools to quantify and determine whether rRNA genes variants differ in their regulation. We made use of single nucleotide polymorphisms that recent works from our and other laboratories have identified at the mouse rRNA promoter (-104) and downstream the transcription start site (+44) (Guetg, Lienemann et al. 2010) (Holland, Lowe et al. 2016). We established SNP-qPCR method that allows specific detection of rRNA gene variants and their quantification. The analysis of two mouse E14 ESC lines, mouse embryonic fibroblast cells (NIH3T3) and mouse C57BL/6-tissue revealed not only a distinct distribution of rRNA variants among individuals but also that not all the variants can be present in the same genome (-104A SNP only in C57BL/6 and NIH3T3 cells, lack of +44T in ESCs+serum). These results clearly suggest that the combination of -104 and +44 SNPs, which are the only SNPs existing at rRNA gene promoter, can be indicative of the presence of several rRNA subclasses. For example, the existence of SNPs at -104 (A or C) and +44 (A or T or G) in C57BL/6-tissue and NIH3T3 cells can give rise to six rRNA gene variants (-104C/+44A, -104C/+44T, -104C/+44G, -104A/+44A, -104A/+44T, -104A/+44G) whereas the ESCs+2i line can have only three variants (-104C/+44A, -104C/+44T, -104C/+44G) and ESCs+serum only two (-104C/+44A, -104C/+44G). Future analyses will be aimed to quantify all these subclasses and determine further variations in regulatory sequences located upstream the main promoter such as the enhancer repeat region, which is also known to differ in length between rRNA copies of the same cells (Santoro, Schmitz et al. 2010). Finally, according to our protocol, the analyses of SNPs at rRNA genes can be easily adapted to human cells. This is of particular importance since rRNA transcription is often altered in cancer ((Nguyen le, Raval et al. 2015). Determine whether alterations of rRNA gene transcription in cancer correlates with changes in rRNA copy number and genes subclasses will be an aim of our future investigation.

The lack of +44T-variants in one of the ESC line (ESCs+serum) was a surprising result since both lines derive from *mus musculus* strain 129/Ola and therefore should have the same genetic background. However, they have been generated in different labs and it can be plausible that ESCs+serum line has lost these sequences during culture. The fact that mouse rRNA genes are clustered together on different 5 chromosomes makes the lack of

+44T-variants an interesting result that is currently under investigation. Indeed, the lack of +44T-sequences can instruct us about the distribution and chromosomal location of rRNA genes, which is so far completely ignored. Indeed, we reason that the lack of +44-rRNA genes in ESCs+serum line can be easily explained if we consider that all +44T-sequences have to cluster together on the same chromosome and that the ESCs+serum line during culture has undergone genomic instability by losing either the chromosome containing the +44T-copies or the entire rRNA locus at one chromosome. We are currently testing this hypothesis by analyzing and comparing the karyotype and the distribution of rRNA genes between the two ESC lines. If this turns to be true (i.e. +44T-variants are all located on one chromosome), this will represent an unprecedented opportunity to track rRNA gene variants in cells and define their spatial localization and organization in the nucleoli.

The establishment of a quantitative method to measure rRNA gene variants allowed exploring how rRNA gene variants are regulated during development. Previous studies have shown that upon ESC differentiation a fraction of rRNA genes is epigenetically silenced through epigenetic modifications mediated by TIP5 (Savic, Bar et al. 2014) (Leone, Bar et al. 2017). An important epigenetic modification at rRNA genes is methylation of DNA, which in mammals occurs at cytosine within the CpG sequences. A question arising from these studies is whether epigenetic silencing at rRNA genes upon exit from pluripotency state is a random event (i.e. any rRNA gene can be epigenetically silenced) or whether a defined set of rRNA genes is specifically CpG methylated and transcriptionally repressed. Our analyses revealed that +44A-rRNA genes are less prone to be methylated whereas +44G and +44T-variants are the major targets for *de novo* methylation. These data strongly indicate that upon ESC differentiation methylation at rRNA genes is not randomly distributed but some rRNA gene variants are more prone to acquire CpG methylation. Similar results were obtained in NIH3T3 cells where the +44A variants are considerably hypomethylated compared to the other genes, a result that is consistent with previous studies (Guetg, Lienemann et al. 2010). Strikingly, -104A-rRNA gene variants in NIH3T3 cells lack completely CpG methylation, supporting a model in which deposition of repressive epigenetic marks is influenced by rRNA sequences.

Previous results have shown that rRNA genes are maintained in a hypomethylated state in ground state pluripotent ESCs (ESCs+2i). The results of this work are consistent with these data and further support a functional link between hypomethylated rRNA genes and the pluripotency state. The analysis of DNA methylation upon ESC differentiation revealed that rRNA genes maintain a hypomethylated state in both ground-state pluripotent cells (ESCs+2i) and ESCs primed toward differentiation (ESCs+serum). These two ESC states can be distinguished by the expression of de novo DNMT3A and DNMT3B (low in ESCs+2i, high in ESCs+serum), which result in distinct CpG methylation content (low in ESCs+2i, high in ESCs+serum) ((Leitch, McEwen et al. 2013),(Ficz, Hore et al. 2013)). Thus, the results shown in this work indicated that also in the case of ESCs with high DNA methylation potential (ESCs+serum), rRNA genes remain refractory to DNA methylation. Recent results have proposed that controlling the epigenetic silencing of rRNA genes is required for the maintenance of pluripotency and the entrance into differentiation pathways (Savic, Bar et al. 2014) (Leone, Bar et al. 2017). One of these studies suggested that an active state of rRNA genes is beneficial for the maintenance of pluripotency since tethering epigenetic silencing at rRNA genes in ESCs+2i primed cells for differentiation through the acquisition of features of differentiated cells such heterochromatic structures, transcription of differentiation genes and impaired formation of teratoma (Savic, Bar et al. 2014). Moreover, a recent work from our group revealed that abrogation of epigenetic silencing at rRNA genes impairs ESC differentiation, suggesting that silent rRNA genes are required to let pluripotency cells to exit from the pluripotent state (Leone, Bar et al. 2017). Although how the epigenetic state of rRNA genes is functionally linked with the pluripotency state of ESCs remains yet to be elucidated, it is clear that transcription of rRNA and ribosome biogenesis are not the unique players of this process. Indeed, one effect observed upon abrogation of epigenetic silencing of rRNA genes during ESC differentiation is cell death (Savic, Bar et al. 2014) (Leone, Bar et al. 2017). It is unlikely that cells with high rRNA levels undergo cell death since ribosome biogenesis is well known to be positively correlated with cell viability and proliferation (Moss 2004). Moreover, the amount of rRNA transcripts does not always depend on the number of active genes since cells with low number of rRNA gene copies can compensate by upregulating elongation rate (Moss, Langlois et al. 2007)). This can also be observed in the transcriptional analysis of this work, which showed that NPCs that

contains about 30% of silent rRNA genes still transcribe at levels similar to ESCs, which have all rRNA genes in an active state. Evidences have started to indicate that silent rRNA genes can contribute to genome organization since induction of rRNA gene silencing in ESCs resulted in a drastic change of genome architecture with the appearance of highly condensed heterochromatic blocks outside the nucleolus, a global increase in the repressive histone mark H3K9me2, maturation of heterochromatin as major and minor satellites and their transcriptional repression as found in differentiated cells (Savic, Bar et al. 2014). Although the mechanism by which the nucleolus acts in the global restructuring of genome architecture is yet unknown, what is clear is that it also depends on the chromatin state of rRNA genes since the gain of heterochromatin at rRNA genes induces the rest of the genome to remodel into highly condensed structures (Leone, Bar et al. 2017). The establishment of a quantitative method to measure rRNA gene variants will help to better understand how the chromatin architecture at defined rRNA genes contribute into this process.

An unexpected result obtained in this study is that the +44T-rRNA gene is the variant with the highest transcription activity in ESCs and is also more prone to acquire de novo methylation during ESC differentiation. Indeed, +44-rRNA genes are ca. 3.5 fold less abundant than +44A-variants, yet the amounts of +44T-transcripts in ESCs are very similar to +44A-rRNA levels. Although the interpretation of these results is based on the assumption that all rRNA genes in ESCs transcribe because they all lack CpG methylation, the results shown in this study are worth of further investigations aimed to understand how specific rRNA genes are targeted for epigenetic silencing upon differentiation. The development of the SNP-qPCR will be an important tool for future chromatin immunoprecipitation (ChIP) analyses to determine the occupancy of the Pol I transcription machinery, the repressor TIP5 as well as histone modifications at rRNA variants before and after differentiation, offering important insights on the composition of chromatin structure.

Taken together, the results of this study represent an initial step toward the deciphering of how genetic variation influences transcription and epigenetic regulation at rRNA genes, which might be useful for the understanding of complex diseases like cancer.

8 References

Altschuler, S. J. and L. F. Wu (2010). "Cellular heterogeneity: do differences make a difference?" Cell **141**(4): 559-563.

Ambrosi, C., M. Manzo and T. Baubec (2017). "Dynamics and Context-Dependent Roles of DNA Methylation." J Mol Biol **429**(10): 1459-1475.

Arnheim, N. and M. Kuehn (1979). "The genetic behaviour of a cloned mouse ribosomal DNA segment mimics mouse ribosomal gene evolution." J Mol Biol **134**(4): 743-763.

Arnheim, N. and E. M. Southern (1977). "Heterogeneity of the ribosomal genes in mice and men." Cell **11**(2): 363-370.

Avilion, A. A., S. K. Nicolis, L. H. Pevny, L. Perez, N. Vivian and R. Lovell-Badge (2003). "Multipotent cell lineages in early mouse development depend on SOX2 function." Genes Dev **17**(1): 126-140.

Azuara, V., P. Perry, S. Sauer, M. Spivakov, H. F. Jorgensen, R. M. John, M. Gouti, M. Casanova, G. Warnes, M. Merkenschlager and A. G. Fisher (2006). "Chromatin signatures of pluripotent cell lines." Nat Cell Biol **8**(5): 532-538.

Baubec, T. and D. Schubeler (2014). "Genomic patterns and context specific interpretation of DNA methylation." Curr Opin Genet Dev **25**: 85-92.

Bestor, T. H. (2000). "The DNA methyltransferases of mammals." Hum Mol Genet **9**(16): 2395-2402.

Bhattacharya, D., S. Talwar, A. Mazumder and G. V. Shivashankar (2009). "Spatio-temporal plasticity in chromatin organization in mouse cell differentiation and during Drosophila embryogenesis." Biophys J **96**(9): 3832-3839.

Bibel, M., J. Richter, K. Schrenk, K. L. Tucker, V. Staiger, M. Korte, M. Goetz and Y. A. Barde (2004). "Differentiation of mouse embryonic stem cells into a defined neuronal lineage." Nat Neurosci **7**(9): 1003-1009.

Bochar, D. A., J. Savard, W. Wang, D. W. Lafleur, P. Moore, J. Cote and R. Shiekhattar (2000). "A family of chromatin remodeling factors related to Williams syndrome transcription factor." Proc Natl Acad Sci U S A **97**(3): 1038-1043.

Bourc'his, D. and T. H. Bestor (2004). "Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L." Nature **431**(7004): 96-99.

Boyer, L. A., T. I. Lee, M. F. Cole, S. E. Johnstone, S. S. Levine, J. P. Zucker, M. G. Guenther, R. M. Kumar, H. L. Murray, R. G. Jenner, D. K. Gifford, D. A. Melton, R. Jaenisch and R. A. Young (2005). "Core transcriptional regulatory circuitry in human embryonic stem cells." Cell **122**(6): 947-956.

Bozhenok, L., P. A. Wade and P. Varga-Weisz (2002). "WSTF-ISWI chromatin remodeling complex targets heterochromatic replication foci." EMBO J **21**(9): 2231-2241.

Bulut-Karslioglu, A., I. A. De La Rosa-Velazquez, F. Ramirez, M. Barenboim, M. Onishi-Seebacher, J. Arand, C. Galan, G. E. Winter, B. Engist, B. Gerle, R. J. O'Sullivan, J. H. Martens, J. Walter, T. Manke, M. Lachner and T. Jenuwein (2014). "Suv39h-dependent H3K9me3 marks intact retrotransposons and silences LINE elements in mouse embryonic stem cells." Mol Cell **55**(2): 277-290.

Campbell, L. L. and K. Polyak (2007). "Breast tumor heterogeneity: cancer stem cells or clonal evolution?" Cell Cycle **6**(19): 2332-2338.

Chambers, I., J. Silva, D. Colby, J. Nichols, B. Nijmeijer, M. Robertson, J. Vrana, K. Jones, L. Grotewold and A. Smith (2007). "Nanog safeguards pluripotency and mediates germline development." Nature **450**(7173): 1230-1234.

Chen, R. Z., U. Pettersson, C. Beard, L. Jackson-Grusby and R. Jaenisch (1998). "DNA hypomethylation leads to elevated mutation rates." Nature **395**(6697): 89-93.

Clark, S. J., J. Harrison and P. L. Molloy (1997). "Sp1 binding is inhibited by (m)Cp(m)CpG methylation." Gene **195**(1): 67-71.

Clos, J., D. Buttgereit and I. Grummt (1986). "A purified transcription factor (TIF-IB) binds to essential sequences of the mouse rDNA promoter." Proc Natl Acad Sci U S A **83**(3): 604-608.

Comai, L., N. Tanese and R. Tjian (1992). "The TATA-binding protein and associated factors are integral components of the RNA polymerase I transcription factor, SL1." Cell **68**(5): 965-976.

Conconi, A., R. M. Widmer, T. Koller and J. M. Sogo (1989). "Two different chromatin structures coexist in ribosomal RNA genes throughout the cell cycle." Cell **57**(5): 753-761.

Cory, S. and J. M. Adams (1977). "A very large repeating unit of mouse DNA containing the 18S, 28S and 5.8S rRNA genes." Cell **11**(4): 795-805.

Defossez, P. A. and I. Stancheva (2011). "Biological functions of methyl-CpG-binding proteins." Prog Mol Biol Transl Sci **101**: 377-398.

Dev, V. G., R. Tantravahi, D. A. Miller and O. J. Miller (1977). "Nucleolus organizers in Mus musculus subspecies and in the RAG mouse cell line." Genetics **86**(2 Pt. 1): 389-398.

Doerks, T., R. R. Copley, J. Schultz, C. P. Ponting and P. Bork (2002). "Systematic identification of novel protein domain families associated with nuclear functions." Genome Res **12**(1): 47-56.

Erickson, J. M. and R. D. Schmickel (1985). "A molecular basis for discrete size variation in human ribosomal DNA." Am J Hum Genet **37**(2): 311-325.

Ficz, G., T. A. Hore, F. Santos, H. J. Lee, W. Dean, J. Arand, F. Krueger, D. Oxley, Y. L. Paul, J. Walter, S. J. Cook, S. Andrews, M. R. Branco and W. Reik (2013). "FGF signaling inhibition in ESCs drives rapid genome-wide demethylation to the epigenetic ground state of pluripotency." Cell Stem Cell **13**(3): 351-359.

Fisher, C. L. and A. G. Fisher (2011). "Chromatin states in pluripotent, differentiated, and reprogrammed cells." Curr Opin Genet Dev **21**(2): 140-146.

Garagna, S., M. Zuccotti, E. Capanna and C. A. Redi (2002). "High-resolution organization of mouse telomeric and pericentromeric DNA." Cytogenet Genome Res **96**(1-4): 125-129.

Garneau, J. E., M. E. Dupuis, M. Villion, D. A. Romero, R. Barrangou, P. Boyaval, C. Fremaux, P. Horvath, A. H. Magadan and S. Moineau (2010). "The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA." Nature **468**(7320): 67-71.

Gaspar-Maia, A., A. Alajem, E. Meshorer and M. Ramalho-Santos (2011). "Open chromatin in pluripotency and reprogramming." Nat Rev Mol Cell Biol **12**(1): 36-47.

Gerber, J. K., E. Gogel, C. Berger, M. Wallisch, F. Muller, I. Grummt and F. Grummt (1997). "Termination of mammalian rDNA replication: polar arrest of replication fork movement by transcription termination factor TTF-I." Cell **90**(3): 559-567.

Gertsenstein, M., L. M. Nutter, T. Reid, M. Pereira, W. L. Stanford, J. Rossant and A. Nagy (2010). "Efficient generation of germ line transmitting chimeras from C57BL/6N ES cells by aggregation with outbred host embryos." PLoS One **5**(6): e11260.

Gifford, C. A., M. J. Ziller, H. Gu, C. Trapnell, J. Donaghey, A. Tsankov, A. K. Shalek, D. R. Kelley, A. A. Shishkin, R. Issner, X. Zhang, M. Coyne, J. L. Fostel, L. Holmes, J. Meldrim, M. Guttman, C. Epstein, H. Park, O. Kohlbacher, J. Rinn, A. Gnirke, E. S. Lander, B. E. Bernstein and A. Meissner (2013). "Transcriptional and epigenetic dynamics during specification of human embryonic stem cells." Cell **153**(5): 1149-1163.

Gonzalez, I. L. and J. E. Sylvester (1995). "Complete sequence of the 43-kb human ribosomal DNA repeat: analysis of the intergenic spacer." Genomics **27**(2): 320-328.

Gorkin, D. U., D. Leung and B. Ren (2014). "The 3D genome in transcriptional regulation and pluripotency." Cell Stem Cell **14**(6): 762-775.

Grozdanov, P., O. Georgiev and L. Karagyozov (2003). "Complete sequence of the 45-kb mouse ribosomal DNA repeat: analysis of the intergenic spacer." Genomics **82**(6): 637-643.

Grummt, I. (2003). "Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus." Genes Dev **17**(14): 1691-1702.

Grummt, I., H. Rosenbauer, I. Niedermeyer, U. Maier and A. Ohrlein (1986). "A repeated 18 bp sequence motif in the mouse rDNA spacer mediates binding of a nuclear factor and transcription termination." Cell **45**(6): 837-846.

Guetg, C., P. Lienemann, V. Sirri, I. Grummt, D. Hernandez-Verdun, M. O. Hottiger, M. Fussenegger and R. Santoro (2010). "The NoRC complex mediates the heterochromatin formation and stability of silent rRNA genes and centromeric repeats." EMBO J **29**(13): 2135-2146.

Guetg, C., F. Scheifele, F. Rosenthal, M. O. Hottiger and R. Santoro (2012). "Inheritance of silent rDNA chromatin is mediated by PARP1 via noncoding RNA." Mol Cell **45**(6): 790-800.

Haltiner, M. M., S. T. Smale and R. Tjian (1986). "Two distinct promoter elements in the human rRNA gene identified by linker scanning mutagenesis." Mol Cell Biol **6**(1): 227-235.

He, S., D. Nakada and S. J. Morrison (2009). "Mechanisms of stem cell self-renewal." Annu Rev Cell Dev Biol **25**: 377-406.

Heix, J., J. C. Zomerdijk, A. Ravanpay, R. Tjian and I. Grummt (1997). "Cloning of murine RNA polymerase I-specific TAF factors: conserved interactions between the subunits of the species-specific transcription initiation factor TIF-IB/SL1." Proc Natl Acad Sci U S A **94**(5): 1733-1738.

Hermann, A., R. Goyal and A. Jeltsch (2004). "The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites." J Biol Chem **279**(46): 48350-48359.

Holland, M. L., R. Lowe, P. W. Caton, C. Gemma, G. Carbajosa, A. F. Danson, A. A. Carpenter, E. Loche, S. E. Ozanne and V. K. Rakyan (2016). "Early-life nutrition modulates the epigenetic state of specific rDNA genetic variants in mice." Science **353**(6298): 495-498.

Ito, T., M. E. Levenstein, D. V. Fyodorov, A. K. Kutach, R. Kobayashi and J. T. Kadonaga (1999). "ACF consists of two subunits, Acf1 and ISWI, that function cooperatively in the ATP-dependent catalysis of chromatin assembly." Genes Dev **13**(12): 1529-1539.

Jackson, J. P., L. Johnson, Z. Jasencakova, X. Zhang, L. PerezBurgos, P. B. Singh, X. Cheng, I. Schubert, T. Jenuwein and S. E. Jacobsen (2004). "Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in Arabidopsis thaliana." Chromosoma **112**(6): 308-315.

Jeltsch, A. (2006). "Molecular enzymology of mammalian DNA methyltransferases." Curr Top Microbiol Immunol **301**: 203-225.

Jones, M. H., N. Hamana, J. Nezu and M. Shimane (2000). "A novel family of bromodomain genes." Genomics **63**(1): 40-45.

Jones, P. A. (1999). "The DNA methylation paradox." Trends Genet **15**(1): 34-37.

Jones, P. A. (2012). "Functions of DNA methylation: islands, start sites, gene bodies and beyond." Nat Rev Genet **13**(7): 484-492.

Jones, P. A. and D. Takai (2001). "The role of DNA methylation in mammalian epigenetics." Science **293**(5532): 1068-1070.

Kuhn, A. and I. Grummt (1992). "Dual role of the nucleolar transcription factor UBF: trans-activator and antirepressor." Proc Natl Acad Sci U S A **89**(16): 7340-7344.

Kurihara, Y., D. S. Suh, H. Suzuki and K. Moriwaki (1994). "Chromosomal locations of Ag-NORs and clusters of ribosomal DNA in laboratory strains of mice." Mamm Genome **5**(4): 225-228.

Lee, J. T. (2012). "Epigenetic regulation by long noncoding RNAs." Science **338**(6113): 1435-1439.

Leitch, H. G., K. R. McEwen, A. Turp, V. Encheva, T. Carroll, N. Grabole, W. Mansfield, B. Nashun, J. G. Knezovich, A. Smith, M. A. Surani and P. Hajkova (2013). "Naive pluripotency is associated with global DNA hypomethylation." Nat Struct Mol Biol **20**(3): 311-316.

Leone, S., D. Bar, C. F. Slabber, D. Dalcher and R. Santoro (2017). "The RNA helicase DHX9 establishes nucleolar heterochromatin, and this activity is required for embryonic stem cell differentiation." EMBO Rep.

LeRoy, G., A. Loyola, W. S. Lane and D. Reinberg (2000). "Purification and characterization of a human factor that assembles and remodels chromatin." J Biol Chem **275**(20): 14787-14790.

Li, E., T. H. Bestor and R. Jaenisch (1992). "Targeted mutation of the DNA methyltransferase gene results in embryonic lethality." Cell **69**(6): 915-926.

Li, J., R. Santoro, K. Koberna and I. Grummt (2005). "The chromatin remodeling complex NoRC controls replication timing of rRNA genes." EMBO J **24**(1): 120-127.

Loh, Y. H., Q. Wu, J. L. Chew, V. B. Vega, W. Zhang, X. Chen, G. Bourque, J. George, B. Leong, J. Liu, K. Y. Wong, K. W. Sung, C. W. Lee, X. D. Zhao, K. P. Chiu, L. Lipovich, V. A. Kuznetsov, P. Robson, L. W. Stanton, C. L. Wei, Y. Ruan, B. Lim and H. H. Ng (2006). "The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells." Nat Genet **38**(4): 431-440.

Lucchini, R. and J. M. Sogo (1992). "Different chromatin structures along the spacers flanking active and inactive Xenopus rRNA genes." Mol Cell Biol **12**(10): 4288-4296.

Marks, H. and H. G. Stunnenberg (2014). "Transcription regulation and chromatin structure in the pluripotent ground state." Biochim Biophys Acta **1839**(3): 129-137.

Martin, G. R. (1981). "Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells." Proc Natl Acad Sci U S A **78**(12): 7634-7638.

Mayer, C., M. Neubert and I. Grummt (2008). "The structure of NoRC-associated RNA is crucial for targeting the chromatin remodelling complex NoRC to the nucleolus." EMBO Rep **9**(8): 774-780.

Mayer, C., K. M. Schmitz, J. Li, I. Grummt and R. Santoro (2006). "Intergenic transcripts regulate the epigenetic state of rRNA genes." Mol Cell **22**(3): 351-361.

McStay, B. and I. Grummt (2008). "The epigenetics of rRNA genes: from molecular to chromosome biology." Annu Rev Cell Dev Biol **24**: 131-157.

Meissner, A. (2010). "Epigenetic modifications in pluripotent and differentiated cells." Nat Biotechnol **28**(10): 1079-1088.

Meshorer, E. and T. Misteli (2006). "Chromatin in pluripotent embryonic stem cells and differentiation." Nat Rev Mol Cell Biol **7**(7): 540-546.

Mitsui, K., Y. Tokuzawa, H. Itoh, K. Segawa, M. Murakami, K. Takahashi, M. Maruyama, M. Maeda and S. Yamanaka (2003). "The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells." Cell **113**(5): 631-642.

Mohn, F. and D. Schubeler (2009). "Genetics and epigenetics: stability and plasticity during cellular differentiation." Trends Genet **25**(3): 129-136.

Morin, R. D., N. A. Johnson, T. M. Severson, A. J. Mungall, J. An, R. Goya, J. E. Paul, M. Boyle, B. W. Woolcock, F. Kuchenbauer, D. Yap, R. K. Humphries, O. L. Griffith, S. Shah, H. Zhu, M. Kimbara, P. Shashkin, J. F. Charlot, M. Tcherpakov, R. Corbett, A. Tam, R. Varhol, D. Smailus, M. Moksa, Y. Zhao, A. Delaney, H. Qian, I. Birol, J. Schein, R. Moore, R. Holt, D. E. Horsman, J. M. Connors, S. Jones, S. Aparicio, M. Hirst, R. D. Gascoyne and M. A. Marra (2010). "Somatic mutations altering EZH2 (Tyr641) in follicular and diffuse large B-cell lymphomas of germinal-center origin." Nat Genet **42**(2): 181-185.

Moss, T. (2004). "At the crossroads of growth control; making ribosomal RNA." Current Opinion in Genetics & Development **14**(2): 210-217.

Moss, T., F. Langlois, T. Gagnon-Kugler and V. Stefanovsky (2007). "A housekeeper with power of attorney: the rRNA genes in ribosome biogenesis." Cell Mol Life Sci **64**(1): 29-49.

Nemeth, A., R. Strohner, I. Grummt and G. Langst (2004). "The chromatin remodeling complex NoRC and TTF-I cooperate in the regulation of the mammalian rRNA genes in vivo." Nucleic Acids Res **32**(14): 4091-4099.

Nguyen le, X. T., A. Raval, J. S. Garcia and B. S. Mitchell (2015). "Regulation of ribosomal gene expression in cancer." J Cell Physiol **230**(6): 1181-1188.

Nichols, J., E. P. Evans and A. G. Smith (1990). "Establishment of germ-line-competent embryonic stem (ES) cells using differentiation inhibiting activity." Development **110**(4): 1341-1348.

Okano, M., D. W. Bell, D. A. Haber and E. Li (1999). "DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development." Cell **99**(3): 247-257.

Olson, M. O. J. (2011). The nucleolus. New York, Springer.

Panov, K. I., J. K. Friedrich, J. Russell and J. C. Zomerdijk (2006). "UBF activates RNA polymerase I transcription by stimulating promoter escape." EMBO J **25**(14): 3310-3322.

Pelletier, G., V. Y. Stefanovsky, M. Faubladier, I. Hirschler-Laszkiewicz, J. Savard, L. I. Rothblum, J. Cote and T. Moss (2000). "Competitive recruitment of CBP and Rb-HDAC regulates UBF acetylation and ribosomal transcription." Mol Cell **6**(5): 1059-1066.

Peng, J. C. and G. H. Karpen (2007). "H3K9 methylation and RNA interference regulate nucleolar organization and repeated DNA stability." Nat Cell Biol **9**(1): 25-35.

Pikaard, C. S., L. K. Pape, S. L. Henderson, K. Ryan, M. H. Paalman, M. A. Lopata, R. H. Reeder and B. Sollner-Webb (1990). "Enhancers for RNA polymerase I in mouse ribosomal DNA." Mol Cell Biol **10**(9): 4816-4825.

Reik, W. (2007). "Stability and flexibility of epigenetic gene regulation in mammalian development." Nature **447**(7143): 425-432.

Rosner, M. H., M. A. Vigano, K. Ozato, P. M. Timmons, F. Poirier, P. W. Rigby and L. M. Staudt (1990). "A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo." Nature **345**(6277): 686-692.

Santoro, R. (2005). "The silence of the ribosomal RNA genes." Cell Mol Life Sci **62**(18): 2067-2079.

Santoro, R. (2011). The nucleolus. The epigenetics of the nucleolus: structure and function of active and silent ribosomal RNA genes The Nucleolus. M. O. J. Olson, Springer. **15:** 57-82.

Santoro, R. and I. Grummt (2001). "Molecular mechanisms mediating methylation-dependent silencing of ribosomal gene transcription." Mol Cell **8**(3): 719-725.

Santoro, R. and I. Grummt (2005). "Epigenetic mechanism of rRNA gene silencing: temporal order of NoRC-mediated histone modification, chromatin remodeling, and DNA methylation." Mol Cell Biol **25**(7): 2539-2546.

Santoro, R., J. Li and I. Grummt (2002). "The nucleolar remodeling complex NoRC mediates heterochromatin formation and silencing of ribosomal gene transcription." Nat Genet **32**(3): 393-396.

Santoro, R., K. M. Schmitz, J. Sandoval and I. Grummt (2010). "Intergenic transcripts originating from a subclass of ribosomal DNA repeats silence ribosomal RNA genes in trans." EMBO Rep **11**(1): 52-58.

Savic, N., D. Bar, S. Leone, S. C. Frommel, F. A. Weber, E. Vollenweider, E. Ferrari, U. Ziegler, A. Kaech, O. Shakhova, P. Cinelli and R. Santoro (2014). "lncRNA maturation to initiate heterochromatin formation in the nucleolus is required for exit from pluripotency in ESCs." Cell Stem Cell **15**(6): 720-734.

Schlesinger, S. and S. P. Goff (2015). "Retroviral transcriptional regulation and embryonic stem cells: war and peace." Mol Cell Biol **35**(5): 770-777.

Scholer, H. R., G. R. Dressler, R. Balling, H. Rohdewohld and P. Gruss (1990). "Oct-4: a germline-specific transcription factor mapping to the mouse t-complex." EMBO J **9**(7): 2185-2195.

Shiao, Y. H., R. M. Leighty, C. Wang, X. Ge, E. B. Crawford, J. M. Spurrier, S. D. McCann, J. R. Fields, L. Fornwald, L. Riffle, C. Driver, O. A. Quinones, R. E. Wilson, K. S. Kasprzak, G. S. Travlos, W. G. Alvord and L. M. Anderson (2011). "Ontogeny-driven rDNA rearrangement, methylation, and transcription, and paternal influence." PLoS One **6**(7): e22266.

Slack, J. M. (2000). "Stem cells in epithelial tissues." Science **287**(5457): 1431-1433.

Smith, A. G., J. K. Heath, D. D. Donaldson, G. G. Wong, J. Moreau, M. Stahl and D. Rogers (1988). "Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides." Nature **336**(6200): 688-690.

Smith, G. P. (1976). "Evolution of repeated DNA sequences by unequal crossover." Science **191**(4227): 528-535.

Smith, Z. D., M. M. Chan, T. S. Mikkelsen, H. Gu, A. Gnirke, A. Regev and A. Meissner (2012). "A unique regulatory phase of DNA methylation in the early mammalian embryo." Nature **484**(7394): 339-344.

Smith, Z. D. and A. Meissner (2013). "DNA methylation: roles in mammalian development." Nat Rev Genet **14**(3): 204-220.

Stancheva, I., R. Lucchini, T. Koller and J. M. Sogo (1997). "Chromatin structure and methylation of rat rRNA genes studied by formaldehyde fixation and psoralen cross-linking." Nucleic Acids Res **25**(9): 1727-1735.

Stefanovsky, V., F. Langlois, T. Gagnon-Kugler, L. I. Rothblum and T. Moss (2006). "Growth factor signaling regulates elongation of RNA polymerase I transcription in mammals via UBF phosphorylation and r-chromatin remodeling." Mol Cell **21**(5): 629-639.

Straight, A. F., W. Shou, G. J. Dowd, C. W. Turck, R. J. Deshaies, A. D. Johnson and D. Moazed (1999). "Net1, a Sir2-associated nucleolar protein required for rDNA silencing and nucleolar integrity." Cell **97**(2): 245-256.

Strohner, R., A. Nemeth, P. Jansa, U. Hofmann-Rohrer, R. Santoro, G. Langst and I. Grummt (2001). "NoRC--a novel member of mammalian ISWI-containing chromatin remodeling machines." EMBO J **20**(17): 4892-4900.

Sylvester, J. E., I. L. Gonzales and E. B. Mougey (2004). Structure and organisation of vertebrate ribosomal DNA. The Nucleolus,. O. MO. New York, Kluwer Acad./Plenum**:** 58-72.

Treangen, T. J. and S. L. Salzberg (2011). "Repetitive DNA and next-generation sequencing: computational challenges and solutions." Nat Rev Genet **13**(1): 36-46.

Tsai, M. C., O. Manor, Y. Wan, N. Mosammaparast, J. K. Wang, F. Lan, Y. Shi, E. Segal and H. Y. Chang (2010). "Long noncoding RNA as modular scaffold of histone modification complexes." Science **329**(5992): 689-693.

Tseng, H., W. Chou, J. Wang, X. Zhang, S. Zhang and R. M. Schultz (2008). "Mouse ribosomal RNA genes contain multiple differentially regulated variants." PLoS One **3**(3): e1843.

Vance, K. W. and C. P. Ponting (2014). "Transcriptional regulatory functions of nuclear long noncoding RNAs." Trends Genet **30**(8): 348-355.

Voigt, P., G. LeRoy, W. J. Drury, 3rd, B. M. Zee, J. Son, D. B. Beck, N. L. Young, B. A. Garcia and D. Reinberg (2012). "Asymmetrically modified nucleosomes." Cell **151**(1): 181-193.

Voigt, P., W. W. Tee and D. Reinberg (2013). "A double take on bivalent promoters." Genes Dev **27**(12): 1318-1338.

Wellauer, P. K., I. B. Dawid, D. D. Brown and R. H. Reeder (1976). "The molecular basis for length heterogeneity in ribosomal DNA from Xenopus laevis." J Mol Biol **105**(4): 461-486.

Williams, R. L., D. J. Hilton, S. Pease, T. A. Willson, C. L. Stewart, D. P. Gearing, E. F. Wagner, D. Metcalf, N. A. Nicola and N. M. Gough (1988). "Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells." Nature **336**(6200): 684-687.

Wu, S. C. and Y. Zhang (2010). "Active DNA demethylation: many roads lead to Rome." Nat Rev Mol Cell Biol **11**(9): 607-620.

Ying, Q. L., J. Wray, J. Nichols, L. Batlle-Morera, B. Doble, J. Woodgett, P. Cohen and A. Smith (2008). "The ground state of embryonic stem cell self-renewal." Nature **453**(7194): 519-523.

Yokochi, T. and K. D. Robertson (2002). "Preferential methylation of unmethylated DNA by Mammalian de novo DNA methyltransferase Dnmt3a." J Biol Chem **277**(14): 11735-11745.

Young, R. A. (2011). "Control of the embryonic stem cell state." Cell **144**(6): 940-954.

Zhou, Y. and I. Grummt (2005). "The PHD finger/bromodomain of NoRC interacts with acetylated histone H4K16 and is sufficient for rDNA silencing." Curr Biol **15**(15): 1434-1438.

Zhou, Y., R. Santoro and I. Grummt (2002). "The chromatin remodeling complex NoRC targets HDAC1 to the ribosomal gene promoter and represses RNA polymerase I transcription." EMBO J **21**(17): 4632-4640.

Zomerdijk, J. C., H. Beckmann, L. Comai and R. Tjian (1994). "Assembly of transcriptionally active RNA polymerase I initiation factor SL1 from recombinant subunits." Science **266**(5193): 2015-2018.

Acknowledgements

First of all I want to thank PD Dr. Raffaella Santoro for giving me the opportunity to write my doctor thesis in her group. I really appreciated the time and energy she spent on me and how supporting and patient she was.

In addition, I thank Prof. Ph.D. Peter Meyer for being my co-reviewer, reading and rating my thesis.

A special thank goes to Dominik Bär and Damian Dalcher who were supporting me during my time in the laboratory. I am thankful for the patience they had answering thousands of my questions and helping me find everything in the eternal depths of freezing drawers and server folders.

Further I want to thank Rodrigo, Sarah, Karolina, Isabella, Christiana and Sergio for the nice atmosphere in the group. Everybody was really helpful and they made me feel welcome from the first day. I really enjoyed working with these people and had a great time.

Last but not least I want to thank my family for supporting me in everything I do and always believing in me!

Curriculum Vitae

Vorname Name	Julia Kuhn
Geburtsdatum	10.09.1991
Geburtsort	Schaffhausen
Nationalität	Schweizerin
Heimatort	Wohlen AG

08/1998 – 06/2004	**Primarschule** (Flurlingen Schweiz)
08/2004 – 06/2010	**Kantonsschule Rychenberg** (Matura, Winterthur, Schweiz)
09/2011 – 08/2016	**Studium der Veterinärmedizin** (Biomedizinische Forschung, Universität Zürich, Zürich, Schweiz)
21.12.2016	**Staatsexamen der Veterinärmedizin** (Universität Zürich, Schweiz)
02/2017- 07/2017	**Dissertation** unter Leitung von: PD Dr. Raffaella Santoro am Institut für Molekulare Mechanismen bei Krankheiten Vetsuisse Fakultät Universität Zürich Direktor: Prof. Dr. Michael O. Hottiger